

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 16:31:11 ; Search time 141.3 Seconds  
(without alignments)  
7375.802 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878  
Sequence: 1 atgagctctcagagctcgtcgc.....ttcacatgaagatgatga 1878

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA:\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfillseq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	87.8	4.7	7218	1	US-08-232-463-14 Sequence 14, Appl
2	75	4.0	272	4	US-09-313-294A-121 Sequence 121, App
3	74.4	4.0	274	4	US-09-313-294A-463 Sequence 463, App
4	58.8	3.1	3489	2	US-08-728-323A-1 Sequence 1, Appl
5	58.8	3.1	3489	4	US-09-298-568-1 Sequence 1, Appl
6	58.8	3.1	3489	4	US-09-410-399-1 Sequence 1, Appl
7	58.8	3.1	32207	2	US-08-770-379-20 Sequence 20, Appl
8	58.8	3.1	32207	4	US-08-757-669A-20 Sequence 20, Appl
9	58.8	3.1	32207	4	US-09-230-371A-20 Sequence 20, Appl
C 10	53.2	2.8	16442	3	US-08-781-891-208 Sequence 208, App
C 11	53.2	2.8	16442	4	US-09-618-166-209 Sequence 209, App
C 12	49.8	2.7	51259	3	US-08-781-891-209 Sequence 209, App
C 13	49.8	2.7	51259	4	US-09-618-166-209 Sequence 209, App
14	47.4	2.5	1276	3	US-09-411-812A-2 Sequence 2, Appl
15	47.4	2.5	1276	3	US-09-411-812A-2 Sequence 2, Appl
16	47.4	2.5	1276	3	US-09-590-113-2 Sequence 2, Appl
17	47	2.5	6617	4	US-09-976-594-268 Sequence 268, App
18	46.6	2.5	2338	1	US-08-425-069-1 Sequence 1, Appl
19	46.6	2.5	2338	2	US-08-317-844B-1 Sequence 1, Appl
20	45.4	2.4	340	5	US-08-182-175A-104 Sequence 104, App
21	45.4	2.4	340	5	PCT-US92-06412-104 Sequence 104, App
22	43	2.3	1926	4	US-09-249-585A-2 Sequence 2, Appl
23	43	2.3	1926	4	US-09-410-399-3 Sequence 3, Appl
24	43	2.3	2580	3	US-09-050-863-2 Sequence 2, Appl
25	43	2.3	2580	4	US-09-359-081-2 Sequence 2, Appl
C 26	43	2.3	5452	2	US-09-130-114-1 Sequence 1, Appl
C 27	43	2.3	8705	4	US-09-647-344A-14 Sequence 14, Appl

28	43	2.3	9600	3	US-08-910-647-1 Sequence 1, Appl
29	43	2.3	9600	4	US-09-620-925-1 Sequence 1, Appl
30	43	2.3	10596	1	US-07-884-811-15 Sequence 15, Appl
31	43	2.3	10596	1	US-07-885-971-15 Sequence 15, Appl
32	43	2.3	10596	1	US-08-087-783A-15 Sequence 15, Appl
33	43	2.3	10596	1	US-08-194-088B-15 Sequence 15, Appl
34	43	2.3	10596	2	US-08-194-087-15 Sequence 15, Appl
35	43	2.3	10596	5	PCT-US93-04648-15 Sequence 15, Appl
C 36	43	2.3	16080	4	US-09-724-566A-48 Sequence 48, Appl
37	42.2	2.2	1251	4	US-09-828-000-1 Sequence 1, Appl
C 38	42.2	2.2	1270	4	US-08-387-805-1 Sequence 1, Appl
39	42.2	2.2	1958	4	US-09-702-327-3 Sequence 3, Appl
40	41.2	2.2	2277	1	US-08-676-967-2 Sequence 2, Appl
41	41.2	2.2	2277	1	US-08-676-974-2 Sequence 2, Appl
42	41.2	2.2	2277	2	US-09-098-487-2 Sequence 2, Appl
43	41.2	2.2	8920	3	US-08-446-855A-1 Sequence 1, Appl
44	41.2	2.2	8920	2	US-09-150-741-1 Sequence 1, Appl
45	41	2.2	966	2	US-08-766-738-2 Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/C  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-9300  
; TELEFAX: (703) 683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F18  
; US-08-232-463-14  
Query Match 4.7%; Score 87.8; DB 1; Length 7218;

Oy	1009	TATGAGGCTCTTAGAGCAGCGCATTCCTATGGTCCACAGGGGCATCGGGGATGAGTGT	1068
Db	2	TATGAAGCAAGTAAAGCAGCGTCATGTATATCTCCCTCTGGGCACCGGTGATAGAGCGT	61
Oy	1069	CTGATCTTTGAGCGCAGTGCACATGGCTATTTTGAGGCGGAACGCGCTCCACCGGAGTTA	1128
Db	62	TTATATTTTGAAGCTCAGCTGTGGGCTTACATGGAAGCGTGAATTAACACTTT	121
Oy	1129	GCTGAGATGGGGTTAGATGAATTGGCTGG--GGTCAGAGCGCAGTATGTTTCTGGA	1188
Db	122	GTTAAACAGGTACAGCAGGAATTCATGGCACCTACGAGAGTTGATTTGTGCTGGT	181

RESULT 4  
 US-08-728-323A-1  
 ; Sequence 1, Application US/08728323A  
 ; Patent No. 5948676  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: Chang, Yuan  
 ;  
 ; APPLICANT: Bohenzky, Roy A.  
 ;  
 ; APPLICANT: Russo, James J.  
 ;  
 ; APPLICANT: Edelman, Isidore S.  
 ;  
 ; APPLICANT: Moore, Patrick S.  
 ;  
 ; TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
 ;  
 ; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA



TITLE OF INVENTION: Encoding Same And Uses Thereof  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSK/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURES:  
NAME/KEY: CDS  
LOCATION: 1..3489  
US-08-728-323A-1

Query Match 3.1%; Score 58.8; DB 2; Length 3489;  
Best Local Similarity 43.4%; Pred. No. 6.7e-07;  
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

QY 1218 GAAGCAAGATCTGGACATATTCATCAACTCTCAAGGCAAAACAGGCTGAATTGCA 1277  
DB 1800 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1859  
QY 1278 GTTGAATCATACCAAGATGTTGTAAGAGCTGAGCAGATCTGAGAGCAATCA 1337  
DB 1860 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1919  
QY 1338 GCAGCTGAATCTTTAAGAACAGCTCTCAAAACAGAACAGCAGCAGAGTCTTGA 1397  
DB 1920 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1979  
QY 1398 GGAATCTCTGGAATTATGAGCAGAGCTGCTGAGAACTGCAAGAGATTAATCGATCGT 1457  
DB 1980 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2039  
QY 1458 GAGACAGAGAACTAAGATGACAGATGAGCAGAACAGGAGAGATGAGCAGCAGAG 1517  
DB 2040 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2099  
QY 1518 GTTTTCATGATTCATCAATCAACAGATCCATGAAAGAGAGCAGCAAGAGAGATT 1577  
DB 2100 GCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2159  
QY 1578 CGAGATGTTGACAGCAGCAGAACTGTCAGAGTTGTTGTCAGCAGCAGCAGCAATTA 1637  
DB 2160 GCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2219  
QY 1638 TCCCTCTAGCATGAGATGAGTGCAGAAAGAGAGCTGAGAGATGAGTCAAGCTTATGAGTT 1697  
DB 2220 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 2279

QY 1698 TCAAGAGAAAGATGAGAGACTTTGTGAGAGAGAGAGATGCTGATTAAGATCAAGA 1757  
DB 2280 GCAGCAGAGAGCAGCAGCAGATGAGAGAGAGCAGAGATGAGAGATCAAGA 2339  
QY 1758 GAAGAGATGAGAGACATGAAAGAGAGATCAAGAGAGATTTGATCTGAGAGAAAGA 1817  
DB 2340 GCAGAGTTAGAGAGCAGCAGCAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGCA 2399  
QY 1818 ATTGATGAGCTTTGAGAACAG 1839  
DB 2400 GCAGCAGAGATTTAGAGAGCAG 2421

RESULT 5  
US-09-298-568-1  
Sequence 1, Application US/09298568  
Patent No. 6322792  
GENERAL INFORMATION:  
APPLICANT: Kieff, Elliott D.  
APPLICANT: Ballester, Mary E.  
APPLICANT: Kave, Kenneth M.  
TITLE OF INVENTION: RHADINO VIRUS IANA ACTS IN TRANS ON A UNIT OF RHADINO  
FILE REFERENCE: 16412-1000IR  
CURRENT APPLICATION NUMBER: US/09/298,568  
EARLIER FILING DATE: 1999-04-21  
EARLIER APPLICATION NUMBER: US 60/109,422  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patencin Ver. 2.0  
LENGTH: 3489  
SEQ ID NO 1  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-298-568-1

Query Match 3.1%; Score 58.8; DB 4; Length 3489;  
Best Local Similarity 43.4%; Pred. No. 6.7e-07;  
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

QY 1218 GAAGCAAGATCTGGACATATTCATCAACTCTCAAGGCAAAACAGGCTGAATTGCA 1277  
DB 1800 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1859  
QY 1278 GTTGAATCATACCAAGATGTTGTAAGAGCTGAGCAGATCTGAGAGCAATCA 1337  
DB 1860 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1919  
QY 1338 GCAGCTGAATCTTTAAGAACAGCTCTCAAAACAGAACAGCAGCAGAGTCTTGA 1397  
DB 1920 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1979  
QY 1398 GGAATCTCTGGAATTATGAGCAGAGCTGCTGAGAACTGCAAGAGATTAATCGATCGT 1457  
DB 1980 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2039  
QY 1458 GAGACAGAGAACTAAGATGACAGATGAGCAGAACAGGAGAGATGAGCAGCAGAG 1517  
DB 2040 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2099  
QY 1518 GTTTTCATGATTCATCAATCAACAGATCCATGAAAGAGAGCAGCAAGAGAGATT 1577  
DB 2100 GCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2159  
QY 1578 CGAGATGTTGACAGCAGCAGAACTGTCAGAGTTGTTGTCAGCAGCAGCAGCAATTA 1637  
DB 2160 GCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2219  
QY 1638 TCCCTCTAGCATGAGATGAGTGCAGAAAGAGAGCTGAGAGATGAGTCAAGCTTATGAGTT 1697  
DB 2220 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2279  
QY 1698 TCAAGAGAAAGATGAGAGATTTGTGAGAGAGAGAGATGCTGATTAAGATCAAGA 1757



QY	1458	GAGCAGAGAACTTAAGATGCAAGCATGAACGAACAGGAAAGATGGATGACACACAG	1517
Db	19957	GCACAGGATGAGCAGACGACGAGTGTGACGACGACGACGATGACACGACGACGATGACA	19898
QY	1518	GTTTTTCATGGATTCAATCAACAGATCCATGAAAGAGAAGCAAGAGAGGAGATTT	1577
Db	19897	GGACGACAGGATGACGAGGAGCAGCAGATGACGACGACGACGACGATGAGCAGCAGCA	19838
QY	1578	CGAATGTTGCAGCAGCAGGAAAGCTGCCAAGTTGTTGGCCAGCAGCAGCAACTTAA	1637
Db	19837	GGATGACACACAGCAGCAGGATGAGCAGCAGCAGCAGATGACACACAGCAGCAGATCA	19778
QY	1638	TCCCTTAGCAATCAGCATTTGCCGAAAGAGAGCGCGGAAGTGTCAAGCTTCAATGAGTT	1697
Db	19777	GCAGCAGCAGCAGATGACGACGACGACGACGATGAACGACGACGACGAGGAGCAGGA	19718
QY	1698	TCAAGAGAAAGAGATGAGGAGATTTGTGGAAGAGAGGAGATGCTGATAAAGATCAAGA	1757
Db	19717	GCACGAGAGGAGGAGGAGGAGGAGTTAAGAGGACGAGCAGAGATTTAGAGATCAAGA	19658
QY	1758	GAAAGAAATGGAAGACATGAAGAAGAGCATCAGAGAGATATTTTGATCTGGAGAAAGA	1817
Db	19657	GCAAGAGTTTGAAGAGCAGGACGAGGATTAGAGGAGCAGGACGACGAGAGTTAGAGAGCA	19598
QY	1818	ATTGATGAGGCTTTGGACAG	1839
Db	19597	GGACGAGAGTTTAGGAGCAG	19576

RESULT 8  
US-08-757-669A-20/c  
; Sequence 20, Application US/08757669A  
; Patent No. 6183751

```

1  GENERAL INFORMATION:
2  APPLICANT: Chang, Yuan
3  APPLICANT: Bohenzky, Roy A.
4  APPLICANT: Russo, James J.
5  APPLICANT: Edelman, Isidore S.
6  APPLICANT: Moore, Patrick S.
7  TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
8  TITLE OF INVENTION: SEQUENCES AND USES THEREOF
9  NUMBER OF SEQUENCES: 20

```

ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.

```

?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/757,669A
?

```

CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-F  
TELECOMMUNICATION INFORMATION:

```

;
; TELEPHONE: (212) 278-0400
;
; TELEFAX: (212) 391-0525
;
; INFORMATION FOR SEQ ID NO: 20:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 32207 base pairs
;

```

TYPE: nucleic acid  
STRANDEDNESS: double  
MONOMER: ?

MOLECULE TYPE: DNA (genomic)  
US-08-757-669A-20

Query Match	3.1%;	Score 58.8;	DB 3;	Length 32207;
Best Local Similarity	43.4%;	Pred. No. 2.5e-06;		
Matches 270;	Conservative	0;	Mismatches 352;	Indels 0;
			Gaps	0;

QY	1218	GAAGCAGAGACTCTGAGCACTATTTCAATCAACACTCTCAAGCCAAACAAAGGCTCAAAATTGCA	1277
Db	20197	GCACGAGATGACACGACGACGAGATGACGACGACGAGATTGACGACGACGACGAGATTGAGCA	20138
QY	1278	GTTGAAATTCATVCCAGAGATGGTTGTAAAGAGCTTGAGCGAGATCTCTGAGACATCA	1337
Db	20137	GCAGCAGATGATGACACGACGAGGATGAGCAGCAGCGATGTAGCAGCAGCAGACGATGAGCA	20078
QY	1338	GCAGCTGAACTTACTTTTAAGAACAAAGCTCTCAAAACGAACCAAGCCGCCAAGTGTCTTA	1397
Db	20077	GCACGACGAGATGACGACGACGAGATGACGACGACGAGATGACGACGACGACGAGATGAGCA	20018
QY	1398	GGAATCTCTGAAATTTATGAGCCAGAGAGCTGCGTAGAACCTGCAAGAGATTAATCGATCGT	1457
Db	20017	GCAGCAGATGATGACGACGACGAGATGAGCAGCAGCAGATGTAGCAGCAGCAGATGAGCA	19958
QY	1458	GAGACGAGAACTTAAGATGACGATGAACGAACGAGGAAGATGATGCAACGACAG	1517
Db	19957	GCAGCAGAGATGACACGACGACGAGATGAGCAGCAGATGTAGCAGCAGCAGCAGCAGATGAGCA	19898
QY	1518	GTTTTTCATGGAATTCATCAATCAACAGATCCATGAAAGAAAGACGCAACGAAGAGAGATTT	1577
Db	19897	GCAGCAGAGATGACGACGAGGACGACGAGATGACGAGGACGACGAGATGACGACGAGCA	19838
QY	1578	CGAGATGTTTCAGCAGCAGAGGAACGTGCCAAAGTTGTTGGCCAGCAGCAGCAGAACATTAA	1637
Db	19837	GGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGATGTAGCAGCAGCAGCAGCAGATGCA	19778
QY	1638	TCCCTCTAGCAATGACGATTTGCCGAAGAAGACGTGAGGAAGTGTCAAGCTTATGAGATT	1697
Db	19777	GCAGCAGCAGCAGAGATGACGACGACGACGAGATGAACGAGCAGCAGCAGAGACGAGCA	19718
QY	1698	TCAAGAGAAAGATGATGAGGAGTTTGTGGAAGAGACGAGATGCTGATTAAGATCAAGA	1757
Db	19717	GCAGCAGAGGACGACGAGAGAGTTAAGAGACGAGGACGAGAGATTGAGAGATCAAGA	19658
QY	1758	GAGAGAGATGAGAGACATGAAAGAGAGGACATCAAGAGGATTTGATCTGAGAAAGA	1817
Db	19657	GCAGGAGTTTGAAGACAGAGAGCAGAGTTTGAAGAGCAGAGCAGAGATTTGAGAGACA	19598
QY	1818	ATTGATGAGGCTTTGGAACAG	1839
Db	19597	GGAGCAGAGATTTAGAGAGCAG	19576

RESULT 9  
US-09-230-371A-20/c  
; Sequence 20, Application US/09230371A

; GENERAL INFORMATION:

APPLICANT: Bohenzky, Roy P

APPLICANT: Edelman, Theodore S

APPLICANT: Moore, Patrick S  
 ITEM OF INTEREST: INTRON 10000

; TITLE OF INVENTION: USES THEREOF

CURRENT APPLICATION NUMBER: US/09/230,371A

PRIOR APPLICATION NUMBER: PCT/US97/13346

PRIOR FILING DATE: 1997-07-22

; SOFTWARE: PatentIn Ver. 2.0

LENGTH: 32207

ORGANISM: Kaposi's sarcoma-associated herpesvirus

US-09-230-371A-20

Query Match	3.1%	Score 58.8;	DB 4;	Length 32207;
Best Local Similarity	43.4%	Pred. No. 2.5e-06;		
Matches 270;	Conservative	0;	Mismatches 352;	Indels 0;
				Gaps 0;

[illegible]

RESULT 10  
 US-08-781-891-208/c  
 ; Sequence 208, Application US/08781891  
 ; Patent No. 6090620  
 ; GENERAL INFORMATION:  
 APPLICANT: Fu, Yang-Hui  
 APPLICANT: Yu, Chang-En  
 APPLICANT: Oshima, Junko  
 APPLICANT: Mulligan, John T.  
 APPLICANT: Schellenberg, Gerald D.  
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
 NUMBER OF INVENTIONS: WERNER'S SYNDROME  
 NUMBER OF SEQUENCES: 209  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED and BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

COMPTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16442 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

Query Match	2.8%;	Score 53.2;	DB 3;	Length 16442;
Best Local Similarity	47.6%;	Pred. No. 7.6e-05;		
Matches 157;	Conservative	0;	Mismatches 173;	Indels 0;
				Gaps 0

[illegible]

RESULT 11  
US-09-618-166-208/c  
; Sequence 208, Application US/09618166  
; Patent No. 6583112  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
Yu, Chang-Bn  
Oshima, Junko  
Mulligan, John T.  
Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDION TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30.00
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/618,166
  FILING DATE: 17-Jul-2000
  CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
  NAME: Mcmasters, David D.
  REGISTRATION NUMBER: 33,963
  REFERENCE/DOCKET NUMBER: 240052.419C1
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (206) 622-4900
  TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 208:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 16442 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-618-166-208

```

Query Match	2.8%;	Score 53.2;	DB 4;	length 16442;
Best Local Similarity	47.6%;	Pred. No. 7.6e-05;		
Matches 157; Conservative	0;	Mismatches 173;	Indels 0;	Gaps 0;

[illegible]

RESULT 12  
 US-08-781-691-209/c  
 : Sequence 209 Application US/08781891  
 : Patent No. 6090620  
 : GENERAL INFORMATION:  
 : APPLICANT: Fu, Ying-Hui  
 : APPLICANT: Yu, Chang-En  
 : APPLICANT: Oshima, Junko  
 : APPLICANT: Mulligan, John T.  
 : APPLICANT: Schellenberg, Gerald D.  
 : TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
 : TITLE OF INVENTION: WERNER'S SYNDROME  
 : NUMBER OF SEQUENCES: 209  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: SEED AND BERRY LLP  
 : STREET: 6300 Columbia Center, 701 Fifth Avenue  
 : CITY: Seattle  
 : STATE: Washington  
 : COUNTRY: USA  
 : ZIP: 98104-7092  
 : COMPUTER READABLE FORM:

1 MEDIUM TYPE: Floppy disk  
 2 COMPUTER: IBM PC compatible  
 3 OPERATING SYSTEM: PC-DOS/MS-DOS  
 4 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 5  
 6 CURRENT APPLICATION DATA:  
 7 APPLICATION NUMBER: US/08/781,891  
 8 FILING DATE: 27-DEC-1996  
 9 CLASSIFICATION: 800  
 10  
 11 ATTORNEY/AGENT INFORMATION:  
 12 NAME: No. 6050620tendburg Ph.D., Carol  
 13  
 14 REGISTRATION NUMBER: 39,317  
 15 REFERENCE/DOCKET NUMBER: 240052.419  
 16  
 17 TELECOMMUNICATION INFORMATION:  
 18 TELEPHONE: (206) 622-4900  
 19  
 20 TELEFAX: (206) 682-6031  
 21  
 22 INFORMATION FOR SEQ ID NO: 209:  
 23  
 24 SEQUENCE CHARACTERISTICS:  
 25 LENGTH: 51259 base pairs  
 26 TYPE: nucleic acid  
 27 STRANDEDNESS: single  
 28  
 29 TOPOLOGY: linear  
 30  
 31 US-08-781-891-209

Query Match	2.7%;	Score 49.8;	DB 3;	length 51259;
Best Local Similarity	50.5%;	Pred. No. 0.0015;		
Matches 146;	Conservative	0;	Mismatches 142;	Indels 1;
			Gaps	1;

[illegible]

RESULT 13  
 US-09-618-166-209/C  
 Sequence 209, Application US/09618166  
 Patent No. 6583112  
 GENERAL INFORMATION:  
 APPLICANT: Pu, Ying-Hui  
 Yu, Chang-En  
 Oshima, Junko  
 Mulligan, John T.  
 Schellenberg, Gerald D.  
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
 WERNER'S SYNDROME  
 NUMBER OF SEQUENCES: 209  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Seed Intellectual Property Law Group  
 STREET: 701 Fifth Avenue, Suite 6300  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30

Query Match	2.5%	Score 47.4	DB 3	length 1276
Best Local Similarity	47.2%	Pred. No. 0.0088		
Matches 144	Conservative	0	Mismatches 161	Indels 0
			Gaps 0	
QY	1516	ACGTTTTCATGATTCATTCATCAACATCATCAAGAAAGAGACCCCAAGAGAGAGAT	1575	
Db	607	ATGTCCTTAAATACATTCGACAGAAAGCATGATGAAGCAGAAAGAGAGAAAG	666	

QY	1516	AGGTTTTCATGATTCATCAATCAAAACGATCCATGAAGAAGACCGAAAGGAGGAAT	1575
Db	607	ATGTCTTAAAGATTATACGCAGAGAAACAAAGAAATGATGAAGCAGAAACACAGAG	666
QY	1576	TTTCGATGTGTCACACACAGAAAGTGCACAGGTGTGTGGCCACAGCAGCAGACAT	1635
Db	667	CCGACACAGCAGACGCGAAGCCGACACAGCAGCAGAAAGCCGACACACACAGAG	726
QY	1636	AATCCCTCAGCAATGACGATTCGCCAAAGAGCTGAGAAAGTCTCAAAGTTCATGAG	1695
Db	727	TATTAAGTACGAAACGAAAGGAGGAAAGAGAGGAGGAGGAGAAAGAGAGG	786
QY	1696	TTTCAAGAGAAAGAGATGAGAGATTGTGGAAGAGAGGAGATGCTGTAAGATCA	1755
Db	787	AGGAAAGAAAGACGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	846
QY	1756	GAGAAAGAGATGGAAGACATGAGAGAGAGCATCACAGAGATATTGATCTCGAGAA	1815
Db	847	GAGAGAGAGAGAAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	906
QY	1816	GAATT 1820	
Db	907	AAATT 911	

Wed Mar 31 08:14:28 2004

us-10-030-829-2.rnt

Page 9

Search completed: March 30, 2004, 03:15:08  
Job time : 154.3 secs

---

**This Page Blank (uspto)**



Result No.	Score	Query			Description		
		Match	Length	ID			
1	435	23.2	2205	12	US-10-4425-114-18351	Sequence 18351, A	
2	400.4	21.3	2218	12	US-10-4424-559-90522	Sequence 90522, A	
3	254.6	13.6	1377	12	US-10-4424-559-91177	Sequence 91177, A	
4	215.6	11.5	1589	12	US-10-4424-559-91176	Sequence 91176, A	
5	197.8	10.5	2890	12	US-10-4424-559-36215	Sequence 73594, A	
6	175.4	9.3	1103	12	US-10-424-559-36215	Sequence 36215, A	
7	169.8	9.0	1375	12	US-10-424-559-73521	Sequence 73521, A	
8	109.6	5.8	951	12	US-10-4424-559-99770	Sequence 99770, A	
9	63.6	3.4	310	9	US-09-294-093B-1296	Sequence 1296, Ap	
10	58.8	3.1	3489	12	US-09-898-273-1	Sequence 1, Appl	
11	58.8	3.1	3489	14	US-10-294-804-1	Sequence 1, Appl	
12	58	3.1	96602	15	US-10-086-117-61	Sequence 61, Appl	
13	53.8	2.9	6473	12	US-10-424-559-29202	Sequence 29202, A	
14	53.2	2.8	16442	15	US-10-374-077-208	Sequence 208, App	
15	52.8	2.8	2479	15	US-10-108-260A-1966	Sequence 1966, App	

C 16	52.6	2.8	628	14	US-10-029-386-12859	Sequence 12259, A
C 17	52.6	2.8	700	9	US-09-864-761-15129	Sequence 17529, A
C 18	52.2	2.8	766	9	US-09-864-761-15608	Sequence 15608, A
C 19	52.2	2.8	1944	9	US-09-864-761-2825	Sequence 2825, Ap
C 20	51.6	2.7	2093	15	US-10-014-047-1666	Sequence 1666, Ap
C 21	51.2	2.7	1223197	15	US-10-027-632-172264	Sequence 172264, A
C 22	51	2.7	174	14	US-10-029-386-14330	Sequence 14330, A
C 23	51	2.7	267	9	US-09-864-761-27984	Sequence 27984, A
C 24	51	2.7	301	9	US-09-864-761-17047	Sequence 17047, A
C 25	51	2.7	471	9	US-09-864-761-2776	Sequence 2776, Ap
C 26	51	2.7	474	9	US-09-864-761-11284	Sequence 11284, A
C 27	51	2.7	549	14	US-10-029-386-625	Sequence 625, Appl
C 28	51	2.7	68233	15	US-10-034-650-31	Sequence 31, Appl
C 29	50.6	2.7	234	9	US-09-864-761-19061	Sequence 19061, A
C 30	50.6	2.7	468	9	US-09-864-761-2326	Sequence 2326, Ap
C 31	50.2	2.7	511	9	US-09-864-761-22817	Sequence 22817, A
C 32	50	2.7	221000	15	US-10-174-014-12	Sequence 12, Appl
C 33	49.8	2.7	51259	15	US-10-374-077-209	Sequence 209, Appl
C 34	49.6	2.6	598	14	US-10-029-386-3466	Sequence 3466, Ap
C 35	49.4	2.6	3809	13	US-10-001-870-68	Sequence 68, Appl
C 36	49.2	2.6	475	9	US-09-864-761-6203	Sequence 6203, Ap
C 37	49.2	2.6	506	14	US-10-029-386-80619	Sequence 20619, A
C 38	49.2	2.6	1168	14	US-10-017-161-2179	Sequence 1879, Appl
C 39	49.2	2.6	1168	15	US-10-292-798-1825	Sequence 1825, Ap
C 40	49	2.6	223	14	US-10-029-386-17312	Sequence 17312, A
C 41	49	2.6	360	9	US-09-864-761-28274	Sequence 28274, A
C 42	49	2.6	450	9	US-09-864-761-11696	Sequence 11696, A
C 43	49	2.6	543	13	US-10-029-386-6512	Sequence 3612, Appl
C 44	48.8	2.6	720	13	US-10-101-487-74	Sequence 74, Appl
C 45	48.8	2.6	720	13	US-10-101-487-76	Sequence 76, Appl

## ALIGNMENTS

```

RESULT 1
US-10-425-114-18351
; Sequence 18351, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 18351
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURES:
; OTHER INFORMATION: Clone ID: LIB3079-030-A8_FLI
US-10-425-114-18351

      23.2%, Score 435; DB 12; Length 2205;
Query Match      Best Local Similarity 59.0%; Pred. No. 1,1e-115;
Matches 826; Conservative 0; Mismatches 555; Indels 18; Gaps 4;

Oy      462  TCTCGACGTGAGCATGATGTGCGATTAATGCTTCTTGAGGAAGAATGATTCGGATGCTTT 521
      |||||
Db      452  TTCTGTCCTTGAGGTGGAGAAATGTTGATGGCAACAAATCATCAGATGATGATGATGA 511
      |||||

Oy      522  GGATGATTTCTGATGACGACCTTGGCAGTGATGATTAATGCTCGATGATGACGCAAAAGAG 581
      |||||
Db      512  TGACGATGATTTTGAGTGAATGATATATCAGTGAATGATCATATATTTCTGACGCAAGTGA 571
      |||||

Oy      582  CCATGATTCACGAAGAAGAAATTAAGTGCTCAAAAAGTTCTTTGGACGCTTGATAGCTT 641

```

Db 572 TTTTAAAGCTGAAAAAGAAAGAAAGTGTGTTAAAGTCTTTGAAAGTATGAAATACATT 631  
 Qy 642 GTGATGAGAGAGATTAATGAAACCAAGAGGAGTGGCATTTGTCCAGTTGTGCAAGCG 701  
 Db 632 GAGTTTGGAAACAAATTAATGAAACAACTCGGCAATGGCATTTGCCGCGCATGCAAAATG 691  
 Qy 702 AACTGTGTCATGAGATTTGGTATTA---CCTGACCTCTACTAGCTCAATGAGAGACAA 758  
 Db 692 ACCGTGGGAGATTGACTGTATCAAGAGGCTGACGCTTTAGATCATGTAGAACAA 751  
 Qy 759 AGAGCTAGAGCGATTAAGCTCCATAGAGATTTGCTGAGATTGTTAGAAAGATCTACA 818  
 Db 752 GGGTTCTACAAAGGTTAACTTCAAGGAAATTTGCTGATGCTGAGAGAGAGCTTTC 811  
 Qy 819 GATGAGAGCGCATCTGTCTATTCCTGTGTGAGATTTAGGGAGTGAAGGTTTGG 878  
 Db 812 TCGAGAGGGAATCTTCAAGTGTACGAGCGGTGAAACATTTGGAAATGAAAGAGCTG-- 869  
 Qy 879 TGAGATGAAAAGATTTATGAAATTTGTCTGCTCCATGCTCATCATGATGATPCTAG 938  
 Db 870 -CAAGAAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 928  
 Qy 939 ACTGATAGAGAGATTAAGATTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 998  
 Db 929 CTGGAAGAAAGATGAGATGATTAAGTGAAGGAGATGAGAAACAAAGAGCTCTTGATTA 988  
 Qy 999 CTTCGACAGTATGAGGCTCTTGAAGACGCGATTCATGATGCTCAAGGCGCATGCTG 1058  
 Db 989 TTTTGAAGATTAAGAAAGTAAAGCATGCTCATGCTCATGCTCATGCTCATGCTCATG 1048  
 Qy 1059 GATGAGTGTCTGATGTTTGAAGAGAGTGCCTGCTATTTGAGGCGCAAGCTTCCA 1118  
 Db 1049 TATGAGGCTGTTAATATTGAAAGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1108  
 Qy 1119 CCGGAGTGTGCTGATGAGTGGGTTAGTAAATTTGCTGGGCTCA---GAAGCGAGTAT 1175  
 Db 1109 TAAACCTTTGTTATCAAGGTACAGACAGAAATTCAGGACCTTACGAGAGTGTGAT 1168  
 Qy 1176 GTTTTGGAGGTGTGCGCACTGATGCTTCTTCTTCAACGAAAGCATGCTGACAT 1235  
 Db 1169 TGTGCTGTGTGGAAAAAGGCACTATATGTTTCTTGAACAAAGAGATATGAGGC 1228  
 Qy 1236 ATTCAATCAACACTCTCAAGGCAAAACAGGCTGAAATTTGAGTTGAAATCATACAGA 1295  
 Db 1229 ATTTAACAAGCATGCTGATGTTAAAGCCGCTGAGATGAGTGAAGTGTCTTACAATGA 1288  
 Qy 1296 GATGTTTAAAGAGCTGAGGAGATCTCTGAGGACATCAGCAGCTGAATCTCTTAA 1355  
 Db 1289 GATGTGTGATTTCAATGAAACAGATGATGAAAGCAACACAACTGAAATTCCTGAA 1348  
 Qy 1356 GAACAAGCTCTCAAAACAGAAACAGACGCGCAAGTGTGTTGAGAAATCTCTGAAATTA 1415  
 Db 1349 GAACAAGATGTTAAGACAGAGCGAGCTTAAGACAGTAAAGATCATTTGGGTGTGT 1408  
 Qy 1416 GAGCAGAGCTGCTGATGAACTGCAAGAGATTAATGAGTCTGAGAGACAGAACTAAGAT 1475  
 Db 1409 TACCAGAGGCTACGAGACATATTGAAAGAAATTAATTTGTTAGAGTAAAGCAAGA 1468  
 Qy 1476 GAGAGATGAACAGACAGGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1535  
 Db 1469 GAGAGATGAGATGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1528  
 Qy 1536 CAACAAGATCATGAAAG 1595  
 Db 1529 TGAAGATTTCAAGAGCCACAGAGAGATTAAGAAAGAGAGAGAGAGAGAGAGAGAG 1588  
 Qy 1596 GGAAGCTGCAAGGTTGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1655  
 Db 1589 AGAGCCCTCAAGAGGCGAGAGC-----TTTGTGTTGATTTCTGGAACATGTAAGA 1639  
 Qy 1656 TTGCGAAGAGAGCTGAGAGAGTGTCAAGCTTCAATGAGTTTCAAGAGAGAGAGAGAG 1715

Db 1640 TCGTACGCTAAGAGAAATATGTCAGAGATTCATGATTCGCAAGTCAAGATGTGCG 1699  
 Qy 1716 GAGATTGTGGAAGAGAGAGAGATGCTGATTAAGATCAAGAGAGAGAGAGAGAT 1775  
 Db 1700 AGAGTTGAAGTTGAGAGAGAGAGATGATCAAGTTCACAGAGACAAAGATGAAAGCT 1759  
 Qy 1776 GAGAGAGAGATCAAGAGAGATTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1835  
 Db 1760 CAGAGAGAGATCAAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1819  
 Qy 1836 ACAGCTCATGATCAAGCAT 1854  
 Db 1820 GGGCTGATGAGAGAGCAT 1838

RESULT 2  
 US-10-424-599-90522/c  
 ; Sequence 90522, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 90522  
 ; LENGTH: 2218  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(2218)  
 ; OTHER INFORMATION: unsure at all n locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_52754C.1  
 ; US-10-424-599-90522

Query Match 21.3%; Score 400.4; DB 12; Length 2218;  
 Best Local Similarity 67.1%; Pred. No. 1.4e-105; Indels 10; Gaps 3;  
 Matches 615; Conservative 0; Mismatches 291;

Qy 455 AGAGTTTCTGACGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 514  
 Db 2153 AAGAGTCCCTATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2094  
 Qy 515 ATGCTTTGATGAT-TCTGATGACGCTTCAAGTATGATGATGATGATGATGATGATGAT 573  
 Db 2093 ATGCTATGAAAGTATCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2034  
 Qy 574 CAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 633  
 Db 2033 CAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1974  
 Qy 634 GATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 693  
 Db 1973 GATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1914  
 Qy 694 CAGAGAGAGCTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 750  
 Db 1913 GATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1854  
 Qy 751 AGAGCAAAAGAGCTGAG 810  
 Db 1853 AAAACAAAG 1794  
 Qy 811 GATCTACAGATGAG 870  
 Db 1793 GAACTGCGAG 1734

QY	472	G-----AGGATGATGTGATATAGCTTTCGAGAGAAAGAAATATATCCGATGCTT	522
Db	149	GACCTTCAAAACAATTATATGTTGTTGATGATGAGGAGAGAGAAATCATGATGATTTG	208
QY	523	GATGATTTGTGATGACGACCTTGCAAGTGAATGATTAGCTCGGATGTGATCAAAAGAC	582
Db	209	GAAGATACGTAGATGATGTGATGATGATGACGATGATTCGTGATGCTAGTCAAAAGAGT	268
QY	583	CATGATACCGAAGACAGATAAGTGTCTCAAAAAGTTCTTTGGCAGCTTGATTAACCTTG	642
Db	269	CATGAGAACGTTAAAGAAAAAGCAAAATGTATTAAGAAATCTTTGAGATTTTGGATGCGTTG	328
QY	643	TGCATCGAGCAGATTAATGAACCAAGAGGACAGTGGCATTTGTCCAGCTGTGCAGAACGA	702
Db	329	ACTGTGGAACAGATTAACGAAACCAAGAAAGACAGTGCATTTGTCCAGCTTCCCAAGGTGT	388
QY	703	CCTGTGTCCTCATGATTTGTTATTAAC--CTGCACCCCTACTAGTCATGCGAGGACAAA	755
Db	389	CCTGGAGCTATTGATTGGTACAGAGGACGTGCAGCCCTTGSTGACACATGCCAAAAAGAG	448
QY	760	GGAGCTAAGCCAGTTAAGCTCCATAGAGAAATTGGCTGAAGTTTAAAGATCTAACG	819
Db	449	GGATTAAGAAAGGGTAAAGATTACAGGAGGCTTGTCTAATCTTTGATGAGGAATTAACG	508
QY	820	ATGAGAGCGCATCTGTCAATTCCTTGTGTGATGATTTATGGCGAGTGAAGGCTTTGGGT	879
Db	509	AAAGGGGTATCTTCAGTAATTCACACTGGGGAAGTATTTGAAAGTGAAGAAAGCTTTGAAA	568
QY	880	GAGATGAAGAAAGATTATGAAAATTGTCTGGCCTCCAAATGTCATCATGATTAAGTACA	939
Db	569	GAAGAG--AAAGGACCATGAGATAGTTGGCCCTCCAAATGTTGTCTATTCAGAAATACAAAG	627
QY	940	CTGATTAAGGACGATTAAGTAAAGTGGCT	968
Db	628	CTTGACGAGATGAATAATTAAGCGGAT	656

```

RESULT 4
US-10-424-599-91176
; Sequence: 91176, Application US/10424599
; Publication NO. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91176
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
US-10-424-599-91176

```

	Query Match	Similarity	Score	215.67	DB 12	Length 1589
	Best Local	Similarity	64.6%	Pred. No. 9, 2e-52		
	Matches	Conservative	0	Mismatches 184	Indels 12	Gaps 2
Qy	352	TATGTAACAACCTTGTGSCACCCACCCTGATCTGCGCCTCTTTGGAAGAGATGG	411			
Db	29	TATGAACCAACTATGTGACTCCCAATCCTGTAATTGGCGCTCACTCGAGCATGGCTGG	88			
Qy	412	AATTGGCAGCGAAGAGAGTTCTGCTAGCACAACAGCTGTGCAAGAGTTTCTGACGTG	471			
Db	89	AATGGCAATCCAGACCTGTGTCATTGACGCATTGTAAAGGAGTGAATTTCAACAGAG	148			
Qy	472	G-----AGATGATGTGGATATATGCTTCTTGAGGAAGAAATGATTTCCGATGCTTTG	522			

Query Match	10.5%	Score 197.8	DB 12	Length 2890
Best Local Similarity	69.4%	Pred. No. 2.1e-46		
Matches 284	Conservative 0	Mismatches 122	Indels 3	Gaps 1
Qy	548	GTGATGATTTATGACTCGAGTGTGAGTCAAAAGGCCATGATCAGAAAGCAGATTAACT	607	
Db	2890	GTGATGATTTATGATTCCTGACTACTAGTCAAAATTAGCCACAGACTGTTAAGAAAGACCAAG	2831	
Qy	608	GGTTCAAAAAGTTCTTTGGCAGCTTGAGATAGCTTGTCATCGACGAGATTAATGAAACCAAC	667	
Db	2830	GGTTTAAAGGATTTCTTTGAGAACTTGGATGGCCCTACTATCGAAAAGATCAATGAACGGG	2771	
Qy	668	AGAGGCGTGGCATTTGTCACGCTTGTCAGAAAGCACTGGTCGCATCGATTGGTATA---	724	
Db	2770	AAAGGCGTGGCACTGACAGCTTGTCAGAGTGTCCGAGTGCTATTGATGATGATACAGAG	2711	
Qy	725	ACCTGCACCCCTACTAGCTCATGCGAGCAAAAAGAGCTAGGCGAGTTAAGTTCACATA	784	

Query Match	9.3%	Score 175.4	DB 12	Length 1103
Best Local Similarity	65.7%	Pred. No. 4e-40		
Matches 285	Conservative	0	Mismatches 147	Indels 2
Qy	1172	GTATGTTTCTGAGGTGTTGCGCCACACTGATGCGTTCCTTGCAACGACAGAGATGTGG	1231	
Db	434	GTGGGCTCCCTGGGGGGAATGGCCAGCTCTATGTGATCATGGCTTACAAAGAAAGATGTGG	375	
Qy	1232	ACATATTCAATCAACACTCTCAAGGCAAAACAAGCTGAAATTGGAATTGAAATCATATCC	1291	
Db	374	ACTTTTCAACAGACATTTCCCAAGGTAAATTGACCTCAATATGACATATGATCATATC	315	
Qy	1292	AAAG-ATGGTTGAAGAGCTGAGCAGATCTCTGAGCAATCAAGCACTGAACTAC	1350	
Db	314	AGAGAAATGTTGTACACCAATTTGGCAATGATAGAGCAACCAAGCTCTCTAT	255	
Qy	1351	TTTAAACAAGCTCTCAAAACAGAACGACGCGAAGTGCTTTAGGAATCTCTGGA	1410	
Db	254	TTTAAACAAGGCTGTGAAAAGTGCACAAACACACAAAAGCTCTTGAAGATCTATTTGTT	195	
Qy	1411	ATTATGAGGAAAGCTGGTGTGAATCTGAGAGATTAATCGATCGTAGACAG-AGAAC	1469	
Db	194	ATATATCTTGAGAGGTGCGAAGACAAATGAGGAAATTCGATTTGTAGGAGAGAGAAC	135	
Qy	1470	TAAATGTCAGCATGAACAGAACAGGGAAGAGTGAATGATGATGATTTTTCATGA	1528	
Db	134	TAAATGCAACATGAAAGAACCAAAAAGATGTATATGCAAGAACATTTCTTCAAGA	75	
Qy	1530	TTCAATCAACAGATCCATGAAGAAGAGACCCCAAGAGAGAAATTTGAGATTTTGA	1589	
Db	74	CCAGATCAAAATCATTCATGATTTCAAGGGATCCAAAGAGAGAAATTTCCAGAGATGC	15	



```

; SEQ ID NO 1296
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Zea mays
;
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700344080H1
; NAME/KEY: unsure
; LOCATION: 309
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-1296

```

Query Match 3.4%; Score 63.6; DB 9; Length 310;

Best Local Similarity 53.3%; Pred. No. 7.6e-08; Matches 163; Conservative 0; Mismatches 134; Indels 9; Gaps 1;

```

QY 1476 GCAGCATGAACAGACGAGAGAGATGATGACACGACAGCTTTTCATGATTCAT 1535
DB 1 GAAGCAGATGAGATGACGAGAGAGATGAGATCCGAGAGAAATTTTCATGATTCAT 60
QY 1536 CAACAGATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1595
DB 61 TGAGAGATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 1596 GGAACGTGCCAAGGTTGTTGCCACAGCAGCAGAGACATTAATCCCTTAGCATGACGA 1655
DB 121 AGAGGGCTCAAGGCGCAGACGCTTTGATG-----TTGATTCGAGAACTATGAAAGA 171
QY 1656 TTCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1715
DB 172 TCGTAGCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 231
QY 1716 GGAAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1775
DB 232 AGAGTTTGAAGTTGAGCAGAGATGAGATGATCAAAAGTCCAGAGAGAGAGAGAGAG 291
QY 1776 GAAGAA 1781
DB 292 CAAGAA 297

```

RESULT 10  
US-09-894-273-1  
; Sequence 1, Application US/09894273  
; Publication No. US20040037847A1  
; GENERAL INFORMATION:  
; APPLICANT: Kieff, Elliott D.  
; APPLICANT: Ballester, Mary E.  
; APPLICANT: Kaye, Kenneth M.  
; TITLE OF INVENTION: RHADINO VIRUS IANA ACTS IN TRANS ON A UNIT OF RHADINO  
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE  
; FILE REFERENCE: 16412-10001R  
; CURRENT APPLICATION NUMBER: US/09/894,273  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 60/109,422  
; PRIOR FILING DATE: 1998-11-19  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 3489  
; TYPE: DNA  
; ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-894-273-1

Query Match 3.1%; Score 58.8; DB 12; Length 3489;

Best Local Similarity 43.4%; Pred. No. 8.2e-06; Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

```

QY 1218 GAAGCAAGATCTGACATTTCAATCACTCTCAAGCAAAACAAGCTGAATTGCA 1277
DB 1800 GCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1859
QY 1278 GTTGAATCATACCAAGAGATGTTGTAAGAGAGCTGAGGAGATCTCTGAGACATCA 1337

```

```

DB 1860 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 1919
QY 1338 GCAGCTGAATCTATTTAAGAAAGCTTCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1397
DB 1920 GCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 1979
QY 1398 GGAATCTCTGAAATTTATGAGCCGAAAGCTGCGTAAGAACTGACAGAGATTAATGAGTCGT 1457
DB 1980 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2039
QY 1458 GAGACAGAGAACTAAGATGACAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1517
DB 2040 GCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2099
QY 1518 GTTTTCATGATTCATCAATCAACAGATCCATGAAGAGAGAGAGAGAGAGAGAGAGAGATTT 1577
DB 2100 GAGCAGCAGATGAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGATGAGCA 2159
QY 1578 CGAGATGTTGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1637
DB 2160 GAGTGAAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGATGA 2219
QY 1638 TCCCTTAGCATGACGATTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1697
DB 2220 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2279
QY 1698 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1757
DB 2280 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2339
QY 1758 GAGAGAGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTTGATCTGAGAGAGAG 1817
DB 2340 GCAAGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTTGAGAGAGAG 2399
QY 1818 ATTGATGAGGCTTTGAGACG 1839
DB 2400 GAGCAGAGATTAAGAGAGAG 2421

```

RESULT 11  
US-10-294-804-1  
; Sequence 1, Application US/10294804  
; Publication No. US20030133948A1  
; GENERAL INFORMATION:  
; APPLICANT: Robertson, Erle S.  
; APPLICANT: Cotter, Murray A.  
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA  
; TITLE OF INVENTION: to Genomic Host DNA  
; FILE REFERENCE: UM-03778  
; CURRENT APPLICATION NUMBER: US/10/294,804  
; CURRENT FILING DATE: 2002-11-14  
; PRIOR APPLICATION NUMBER: US/09/410,399  
; PRIOR FILING DATE: 1999-10-01  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 3489  
; TYPE: DNA  
; ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-10-294-804-1

Query Match 3.1%; Score 58.8; DB 14; Length 3489;

Best Local Similarity 43.4%; Pred. No. 8.2e-06; Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

```

QY 1218 GAAGCAAGATCTGACATTTCAATCACTCTCAAGCAAAACAAGCTGAATTGCA 1277
DB 1800 GCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1859
QY 1278 GTTGAATCATACCAAGAGATGTTGTAAGAGAGCTGAGGAGATCTCTGAGACATCA 1337
DB 1860 GCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1919

```



QY	1338	GCAGCTGAACACTTTAAGAACCAAGCTCTCAAAACAGAACAGCCGCAAGGGTGGTTGA	1397
Db	1920	GCACGACGGAATGACGACGACGACGAGATGACGACGACGACGAGATGACGACGACGATGAGCA	1979
QY	1398	GGAAATCTCTGGAAAATTATGACCGAAGACTCGTAGAATCTGACAGGATTAATCCGATCGT	1457
Db	1980	GCACGAGGATGAGCGACGACGAGATGAGCAGCAGCAGCAGCAGATGAGCAGCAGATGAGCA	2039
QY	1458	GAGACAGAGAACTTAAGATGCAGCATGAACAGAACGGAAGAGATGGATGCACACGACAG	1517
Db	2040	GCACGACGAGTAGCAGCAGCAGCAGATGAGCAACACACGATGAGCAGCAGCAGATGAGCA	2099
QY	1518	GTTTTTCATGATTCATCAATCAACAGATCCATGAAAAGAAAGAACCCMAAGGAGGAATTT	1577
Db	2100	GGACGACGACGATGACGACGACGACGACAGATGACACGAGCAGCAGGATGCGACGACCA	2159
QY	1578	CGAGATGTTGAGCAGCAGCAACCGTGGCCAAAGTGTGGCCACGACGACGCAAGCAATTAA	1637
Db	2160	GGATGAGCAGCAGCAGCAGATGACGACGACGACGACGACGATGAGCAGCAGCAGCAGATGA	2219
QY	1638	TCCCTTAGCANTGACGATGCCCAAAAGAGAGCTGAGGAAGTGTCAAGCTTCATCGAGTT	1697
Db	2220	GCACGACGACGAGATGACGACGACGACGACGACGACGATGAAACGACGACGACGAGACGAGA	2279
QY	1698	TCAAAGAAAAGATGAGAGAGTTTGTGAAGAGAGGAGAGATGCTGTAAAGATCAGA	1757
Db	2280	GCACGACGAGAGCAGGACGAGGATTAAGAGACAGAGACGAGATTAAGAGATCAGA	2339
QY	1758	GAAACAAGATGGAACATGAAGAAAGGAGCATCAGAGAGATATTGATCTCGAAGAAAGA	1817
Db	2340	GCACGAGATTGAAGAGCAGGAGCAGGAGTTAGAGAGACGAGACGAGAGTTAGAGAGCA	2399
QY	1818	ATTGATGAGGCTTTGGAACAG	1839
Db	2400	GGACGACGAGTTAAGAGAGCAG	2421

```

RESULT 12
US-10-085-117-61/c
: Sequence 61, Application US/10085117
: Publication No. US20030232334A1
: GENERAL INFORMATION:
:   APPLICANT: Morris, David W.
:   TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
:   FILE REFERENCE: 529455000121
:   CURRENT APPLICATION NUMBER: US/10/085,117
:   CURRENT FILING DATE: 2002-02-27
:   PRIOR APPLICATION NUMBER: US 09/798,586
:   PRIOR FILING DATE: 2001-03-02
:   NUMBER OF SEQ ID NOS: 361
:   SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 61
:   LENGTH: 96602
:   TYPE: DNA
:   ORGANISM: Mus musculus
:   FEATURE:
:     NAME/KEY: variation
:     LOCATION: (1)...(96602)
:     OTHER INFORMATION: n = any nucleotide
: US-10-085-117-61

```

Query Match	3.1%	Score 58;	DB 15;	Length 96602;
Best Local Similarity	44.4%;	Pred. No. 0.0001;		
Matches 189;	Conservative	0;	Mismatches 236;	Indels 1;
				Gaps 1;

[illegible][illegible]

```

RESULT 13
US-10-424-599-29202/c
Sequence 29202, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ. ID NOS: 285684
SEQ ID NO 29202
LENGTH: 473
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_12636C.1
US-10-424-599-29202

```

Query Match	2.9%	Score 53.8;	DB 12;	Length 473;
Best Local Similarity	53.6%;	Pred. No. 7.1e-05;		
Matches 112; Conservative	0;	Mismatches 97;	Indels 0;	Gaps 0;

Qy	1650	TGACGATTTGCCAAAAGAGCGTGAAGAAGTGCAAGCTTCATCAGATTCAAGAGAAAG	1705
Db	452	TGAGGATATAGACTCAAGGTGGAAAGTATCTGAAATTTGTGAGTCCAGAAATTGGG	393
Qy	1710	GATGAGAGAGTTTGTGAMAGAGGAGATGCTGATTAAGATCAAGAGAAAGATGGA	1765
Db	392	CTGAGAGAACTTTGTGCTGAAAAAGAGAAAGCTTGCACTCATGGCAATGGGTTATGC	333
Qy	1770	AGACATGAAAGAGGCATCACGAGAGATTTTGATCTGGAGAAAGATTTGATGAGCC	1822
Db	332	TGCAATGAAACGAGGCATTGGGAGAGAAAGTTCAATATGACAGAAAGTTTGATGAGGA	273
Qy	1830	TTTGGAAACAGCTCATGTACAAGCATGGCC	1858
Db	272	ACTGGCCAACCTCATGAGAAAAGTACTCC	244

RESULT 14  
US-10-374-077-208/c

Sequence 208, Application US/10374077  
Publication No. US20040006779A1  
GENERAL INFORMATION:  
APPLICANT: Fu, Yang-Hui  
Yu, Chang-En  
Oshima, Junko  
Mulligan, John T.  
Schellenberg, Gerald D.  
TITLE OF INVENTION: ANTIBODIES AGAINST GENE PRODUCTS RELATED TO  
WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/374,077  
FILING DATE: 25-Feb-2003  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Roseman, Stephen  
REGISTRATION NUMBER: 43,058  
REFERENCE/DOCKET NUMBER: 100107.401D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16442 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 208:  
US-10-374-077-208  
Query Match 2.8%; Score 53.2; DB 15; Length 16442;  
Best Local Similarity 47.6%; Pred. No. 0.0009;  
Matches 157; Conservative 0; Mismatches 173; Indels 0; Gaps 0;  
QY 1472 AGATGCAAGCATGAACAGAACAGGAGAGAGATGATGACACAGAGTTTTCATGATT 1531  
DB 16437 AGGAG 16378  
QY 1532 CAATCAACAGATCCATGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1591  
DB 16377 AGCAG 16318  
QY 1592 ACCAG 16258  
DB 16317 AGGAG 1711  
QY 1652 AGCATTTGCGAAG 16198  
DB 16257 AGGAG 1771  
QY 1712 TGGAG 16138  
DB 16197 AGGAG 1801  
QY 1772 ACATGAAG 16108  
DB 16137 AGAAG

RESULT 15

US-10-108-260A-1966  
Sequence 1966, Application US/10108260A  
Publication No. US20040005560A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20040005560A1 full length cDNA  
FILE REFERENCE: H1-A0106  
CURRENT APPLICATION NUMBER: US/10/108,260A  
CURRENT FILING DATE: 2002-03-27  
NUMBER OF SEQ ID NOS: 5458  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1966  
LENGTH: 2479  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-108-260A-1966  
Query Match 2.8%; Score 52.8; DB 15; Length 2479;  
Best Local Similarity 43.9%; Pred. No. 0.0038;  
Matches 225; Conservative 0; Mismatches 287; Indels 0; Gaps 0;  
QY 1328 AGGACATATGACGAGCTGACTCTTAAGAACAAAGCTCTCAAAACAGAACAGAGCCCA 1387  
DB 1377 AGGAGAAAAGATGTGGAGGAGAGAGAGAGAGAGATTCAGACAGAGAGAGATACGGAGG 1436  
QY 1388 AGGTGCTTGAAGATCTCGAAATTTATGACGAGAGAGCTGTAAGACTGACAGAGATA 1447  
DB 1437 AGGAGAAAGAGGACGAGAGAGAGAGAGAGATGTGAGGAGAGAGAGAGAGAGAGAG 1496  
QY 1448 ATCCGATCTGAGAGAGAGAGAGAGAGATTAAGATCCAGATGAACAGAACAGAGAGATG 1507  
DB 1497 AGGAGATATGAGAGAGAGAGAGAGAGATGCAAGAGAGAGAGAGATATGAGAGAGAG 1556  
QY 1508 CACACAGAGGTTTTTCATGATTCATCAATCAATCAATCAATCAATCAATCAATCAATCA 1567  
DB 1557 AGAGGTGTGAG 1616  
QY 1568 AGGAGATTTTGGAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1627  
DB 1617 AGGAGAGATGTGAG 1676  
QY 1628 AGACATTAATCCCTTAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1687  
DB 1677 AGGAGAGAGAGATATCAG 1736  
QY 1688 TCATGAGTTTAAAG 1747  
DB 1737 AGCAGAGAGAGAGATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1796  
QY 1748 AAGATCAAG 1807  
DB 1797 GGGAG 1856  
QY 1808 TGGAGAGAGAGATTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1839  
DB 1857 TGGAG 1888  
Search completed: March 30, 2004, 08:58:17  
Job time : 679.885 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 13:41:30 ; Search time 655.116 Seconds  
(without alignments)  
12178.168 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878  
Sequence: 1 atgagctctagggctgctcc.....ttcacatgaagatgatga 1878

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1878	100.0	1878	AAE5374	AAE5374 Nucleotide
2	1462	77.8	3275	AAE5373	AAE5373 Genomic S
3	75	4.0	272	ABX8161	ABX8161 Corn ear-
4	74.4	4.0	274	ABX8203	ABX8203 Corn ear-
5	63.6	3.4	310	ABL71922	ABL71922 Corn tass
6	58.8	3.1	3489	AAA0290	AAA0290 Kaposi's
7	58.8	3.1	3489	AAE82901	AAE82901 Nucleotide
8	58.8	3.1	3489	ABA93487	ABA93487 Kaposi's
9	58.8	3.1	32207	AAV73805	AAV73805 KSHV LTR
10	58.8	3.1	137507	AAV19941	AAV19941 KSHV long
11	56.2	3.0	1080	AAI99537	AAI99537 Mouse lnc
12	55	2.9	1686	AAQ87587	AAQ87587 DNA encod
13	55	2.9	4000	AAE91902	AAE91902 Mammos-1
14	53.2	2.8	1140	AAE40071	AAE40071 Arabidops
15	53.2	2.8	16442	AAE83006	AAE83006 Partial m
16	53	2.8	3399	AAE05868	AAE05868 Chicken l
17	52.6	2.8	700	AAI20076	AAI20076 Probe #10
18	52.6	2.8	700	ABA65105	ABA65105 Human foe
19	52.6	2.8	700	AAI45274	AAI45274 Probe #13
20	52.6	2.8	700	ABA47220	ABA47220 Human bre
21	52.6	2.8	700	ABA32209	ABA32209 Probe #10
22	52.6	2.8	700	AAK39264	AAK39264 Human bon
23	52.6	2.8	700	AAK13527	AAK13527 Human bra

24	52.6	2.8	700	4	ABG38852	ABg38852 Human liv
25	52.6	2.8	700	5	AAI05782	AAI05782 Probe #57
26	52.6	2.8	700	6	ABG13350	ABs13350 Human gen
27	52.2	2.8	2000	7	ADA71938	ADA71938 Rice gene
28	52	2.8	766	4	AAI22114	AAI22114 Probe #12
29	52	2.8	766	4	ABA67193	ABA67193 Human foe
30	52	2.8	766	4	AAI47409	AAI47409 Probe #16
31	52	2.8	766	4	ABA49279	ABA49279 Human bre
32	52	2.8	766	4	ABA34288	ABA34288 Probe #12
33	52	2.8	766	5	AAI07812	AAI07812 Probe #78
34	52	2.8	1944	4	AAI12874	AAI12874 Probe #28
35	52	2.8	1944	4	ABA54575	ABA54575 Human foe
36	52	2.8	1944	4	AAI34231	AAI34231 Probe #29
37	52	2.8	1944	4	ABA44123	ABA44123 Human bre
38	52	2.8	1944	4	ABA24359	ABA24359 Probe #28
39	52	2.8	1944	5	AAI02791	AAI02791 Probe #27
40	51.6	2.7	2093	9	ADB63512	ADB63512 Human CDN
41	51.2	2.7	693	5	AAE74240	AAE74240 DNA encod
42	51.2	2.7	693	5	AAE90715	AAE90715 DNA encod
43	51	2.7	267	4	AAK45604	AAK45604 Human bon
44	51	2.7	267	4	AAK19599	AAK19599 Human bra
45	51	2.7	267	4	ABG45294	ABg45294 Human liv

#### ALIGNMENTS

RESULT 1	AAE5374	standard; cDNA; 1878 BP.
ID	AAE5374	
AC	AAE5374;	
XX		
DT	15-MAY-2001	(first entry)
DE	Nucleotide sequence of the Arabidopsis SGS3 polypeptide.	
XX		
KW	SGS3 gene; post-transcriptional inactivation; RNA degradation;	
XX	virul resistance; resistance; fatty acid content; protein content; ss.	
OS	Arabidopsis thaliana.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1878
FT		/*tag= a
FT		/product= "SGS3"
XX		
PN	WO200105951-A2.	
XX		
PD	25-JAN-2001.	
XX		
PF	13-JUL-2000; 2000WO-FR002052.	
XX		
PR	16-JUL-1999; 99FR-00009417.	
XX		
PR	26-JAN-2000; 2000FR-00001006.	
XX		
PA	(AVET ) AVENTIS CROSCIENCE SA.	
XX	(INRG ) INST NAT RECH AGRONOMIQUE.	
XX		
PI	Beclin C, Elmayer T, Vaucheret H;	
XX		
DR	WPI; 2001-159529/16.	
XX	P-PSDB; AAB31798.	
XX		
PT	New SGS3 gene from Arabidopsis thaliana, useful for increasing virus	
XX	resistance in plants and, when inhibited, for increasing transgene	
XX	expression.	
PS	Claim 1; Page 32-35; 36pp; French.	
XX		
CC	The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.	
XX	The SGS3 gene is essential for post-transcriptional inactivation	
CC	(degradation of RNA) and for resistance to viruses. Overexpression of	

CC 5GS3 results in plants with increased resistance to viruses, while  
 CC inactivation of 5GS3 in transgenic plants (e.g. by expressing antisense  
 CC RNA, by mutation or by homologous recombination) increases the level of  
 CC the transgene product. This product may e.g. impart resistance (to  
 CC herbicide, insects or pathogens), alter contents of essential fatty acids  
 CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or  
 CC interferon

XX Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 100.0%; Score 1878; DB 4; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAGTTCTAGGGCTGTCCTCAATGCTTAAAGAAAAGAACGTTCAAGGTCGTTAAGGCTT 60
DB 1 ATGAGTTCTAGGGCTGTCCTCAATGCTTAAAGAAAAGAACGTTCAAGGTCGTTAAGGCTT 60
QY 61 GAGGTGAACAGTTGGTTCAAGTTTGGACAGGAGACGACTGGCTTCTTCAAGATGAT 120
DB 61 GAGGTGAACAGTTGGTTCAAGTTTGGACAGGAGACGACTGGCTTCTTCAAGATGAT 120
QY 121 GAGGAGAGTGGAGGTCATTTCAGAGAGAACAGAACAAACAGAACTTCTGGA 180
DB 121 GAGGAGAGTGGAGGTCATTTCAGAGAGAACAGAACAAACAGAACTTCTGGA 180
QY 181 AAAACCTTGGTTCTCGAATTCGATTCCTTGAAGCTTGGGGTGTGACAGCAAGG 240
DB 181 AAAACCTTGGTTCTCGAATTCGATTCCTTGAAGCTTGGGGTGTGACAGCAAGG 240
QY 241 AGAGGTGACACGATCTGGAGAGAGAAACATGATCCGGAGAGGTAAACGCAATGAT 300
DB 241 AGAGGTGACACGATCTGGAGAGAGAAACATGATCCGGAGAGGTAAACGCAATGAT 300
QY 301 CGGGGCAATTCAGCTAACATCTCTGTCGGGAGAGAGGTTGAGCAGAAAGTATGATAC 360
DB 301 CGGGGCAATTCAGCTAACATCTCTGTCGGGAGAGAGGTTGAGCAGAAAGTATGATAC 360
QY 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTGGAGAGAGATGGAATTGGCAG 420
DB 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTGGAGAGAGATGGAATTGGCAG 420
QY 421 GCAAGAGAGGTTCTGCTCAGACACAGCTGTGCGAGGTTCCGACGTGAGATGAT 480
DB 421 GCAAGAGAGGTTCTGCTCAGACACAGCTGTGCGAGGTTCCGACGTGAGATGAT 480
QY 481 GTGCAATATGCTTGAAGAGAGATGATCCGATGCTTGGATGATTTGATGACAC 540
DB 481 GTGCAATATGCTTGAAGAGAGATGATCCGATGCTTGGATGATTTGATGACAC 540
QY 541 CTTCGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 541 CTTCGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 601 AATTAAGGTTCGAAAGGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGAT 660
DB 601 AATTAAGGTTCGAAAGGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGAT 660
QY 661 GAACCAAGAGAGGAGTGGCATTTGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 GAACCAAGAGAGGAGTGGCATTTGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 TATAACCTGCACTCTTAAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TATAACCTGCACTCTTAAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGATCTGCATT 840
DB 781 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGATCTGCATT 840
QY 841 CTTGTGTGATGATTTATGGCAGATGGAAGGTTTGGGTGAGAGTGAAGATTTATGAA 900
DB 841 CTTGTGTGATGATTTATGGCAGATGGAAGGTTTGGGTGAGAGTGAAGATTTATGAA 900

```

```

QY 901 ATTGTGCGCTCCAAATGCTATCATCATGATATCTAGATGATTAAGAGCATTAACGAT 960
DB 901 ATTGTGCGCTCCAAATGCTATCATCATGATATCTAGATGATTAAGAGCATTAACGAT 960
QY 961 AAGTGGCTGGGCAATGGGCAACCAAGAGCTGTGAAATCTTCAACAATATGAGGCTTT 1020
DB 961 AAGTGGCTGGGCAATGGGCAACCAAGAGCTGTGAAATCTTCAACAATATGAGGCTTT 1020
QY 1021 AAGAGCAGCATTTCTATGATGTCACAGGCGCATGTGATGATGATGATGATGATGAT 1080
DB 1021 AAGAGCAGCATTTCTATGATGTCACAGGCGCATGTGATGATGATGATGATGATGAT 1080
QY 1081 AGCAGTCCACAGGCTATTTGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 AGCAGTCCACAGGCTATTTGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 TATGCTTCTTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 TATGCTTCTTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 ACAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 ACAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 ATCTCTGAGACATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 ATCTCTGAGACATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 CACGCCAAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 CACGCCAAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 GAGGATTAATCGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 GAGGATTAATCGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 ATGATGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1501 ATGATGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TCAAGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
DB 1681 TCAAGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
QY 1741 CTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 CTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 TTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1801 TTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 CACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1878
DB 1861 CACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1878

```

RESULT 2  
 AAF25373 standard; DNA; 3275 BP.  
 ID AAF25373



QY 1068 TCTGATGTTTGAAGACAGTGCCTGCTATTTGGAGGCGGAGCCCTCAACGGAGATT 1127  
 DB 1836 TCTGATGTTTGAAGACAGTGCCTGCTATTTGGAGGCGGAGCCCTCAACGGAGATT 1895  
 QY 1128 AGCTGAGATGGGGTATAGATAGATTGCTGGGGTCAAGAGCGCATATGTTTCTGGAGG 1187  
 DB 1896 AGCTGAGATGGGGTATAGATAGATTGCTGGGGTCAAGAGCGCATATGTTTCTGGAGG 1955  
 QY 1188 TGTTCGCCAATCTGTATGCTCTTCTTGGCAAGCAAGCAATCTGGCAATATTCATCAACA 1247  
 DB 1956 TGTTCGCCAATCTGTATGCTCTTCTTGGCAAGCAAGCAATCTGGCAATATTCATCAACA 2015  
 QY 1248 CTCTCA----- 1253  
 DB 2016 CTCTCAAGGTTCTCTCCCCCAAGAAATTTGATATATGCTTTAGTTTGTGATTGAAT 2075  
 QY 1254 -----AGG 1256  
 DB 2076 TTTAAGTTTGTGTCGGTATATGATCTGTATATATCTATGATTCATTAGG 2135  
 QY 1257 CAATAAGAGCTTAATTTGAGTTGAATCATACCAAGAGATGTTGTAAGAGCTGAG 1316  
 DB 2136 CAATAAGAGCTTAATTTGAGTTGAATCATACCAAGAGATGTTGTAAGAGCTGAG 2195  
 QY 1317 GCAGATCTGAGGACATCAGCAGCTGAATCTTTAAGAACACAGCTCTCAAAACAGAA 1376  
 DB 2196 GCAGATCTGAGGACATCAGCAGCTGAATCTTTAAGAACACAGCTCTCAAAACAGAA 2255  
 QY 1377 CAGCAGCGCAAGGCTTGAAGAAATCTTGAATTTATAGGAGAAAGCTGTAGAAC 1436  
 DB 2256 CAGCAGCGCAAGGCTTGAAGAAATCTTGAATTTATAGGAGAAAGCTGTAGAAC 2315  
 QY 1437 TGCAGAGATATTCGATTCGAGAGAGAGAACTAAGATCAGCATGACAGAAACAGGAA 1496  
 DB 2316 TGCAGAGATATTCGATTCGAGAGAGAGAACTAAGATCAGCATGACAGAAACAGGAA 2375  
 QY 1497 AG----- 1498  
 DB 2376 AGAGATGATTTTCTTCAAGAAATCAAACTTGACATTTGTATTAATCTACTGATTC 2435  
 QY 1499 -----AGATGATGCACACGA 1514  
 DB 2436 CATTTTGATTATATTCCTCAACAAACAAACCTGTGGTGGTTGAAGATGATGACACGA 2495  
 QY 1515 CAGGTTTTCATGATCAATCAACAGATCCATGAAGAGAGAGAGAGAGAGAGAGAG 1574  
 DB 2496 CAGGTTTTCATGATCAATCAACAGATCCATGAAGAGAGAGAGAGAGAGAGAGAG 2555  
 QY 1575 TTTGAGAGATTTGAGCAGCAGAGAAAGCTGCAAGGTTTGGCCAGAGAGAGAGAGAG 1634  
 DB 2556 TTTGAGAGATTTGAGCAGCAGAGAAAGCTGCAAGGTTTGGCCAGAGAGAGAGAGAG 2615  
 QY 1635 TTAATCCTCTACCATGAGATGAGGAGAG----- 1665  
 DB 2616 TTAATCCTCTACCATGAGATGAGGAGAGAGATGATGATGATGATGATGATGATG 2675  
 QY 1666 ----- 1665  
 DB 2676 CTGGCGTTTGTGTTTCAACCTAAGATGATGATGATGATGATGATGATGATGATG 2735  
 QY 1666 -AGAGTTGAGAAAGTCAAGCTTCAATGAGTTTCAAGAGAGAGAGAGAGAGAGAG 1724  
 DB 2736 CAGAGCTGAGAAAGTCAAGCTTCAATGAGTTTCAAGAGAGAGAGAGAGAGAGAG 2795  
 QY 1725 GGAAGAGAGAGAGATGCTGATTAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1784  
 DB 2796 GGAAGAGAGAGAGATGCTGATTAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2855  
 QY 1785 GCATCAAGAGAGATTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844  
 DB 2856 GCATCAAGAGAGATTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2915

QY 1845 GTTACAGATGGCTTTCACATGATGATGATGA 1878  
 DB 2916 GTTACAGATGGCTTTCACATGATGATGATGA 2949  
 RESULT 3  
 ABX81661  
 ID ABX81661 standard; cDNA; 272 BP.  
 XX  
 AC ABX81661;  
 XX  
 DE 24-Apr-2003 (first entry)  
 XX  
 DE Corn ear-derived polynucleotide (cpd) #121.  
 XX  
 KW Corn ear-derived polynucleotide; cpd; cDNA library; SATMON022; SATMON023;  
 KW structural gene; functional gene; regulatory gene; gene expression;  
 KW corn ear-specific profile; gene transcription; plant breeding program;  
 KW hybrid plant; desirable trait expression; growth; development;  
 KW inheritance; desired characteristic; quality; yield;  
 KW disease resistance; environmental adaptability; quality; yield;  
 KW multigene trait; plant; gene; se.  
 KW  
 OS Zea mays.  
 XX  
 PN US6476212-B1.  
 PD 05-NOV-2002.  
 XX  
 PF 14-MAY-1999; 99US-00313294.  
 XX  
 PR 26-MAY-1998; 98US-0086722P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Laligudi RV, Ito LY, Sherman BK;  
 XX  
 DR WPI; 2003-208840/20.  
 XX  
 PT Novel purified corn-ear derived polynucleotide useful as hybridization  
 PT probe for detecting polynucleotide in sample, and for identifying,  
 PT evaluating, and altering desired characteristics associated with growth,  
 PT development.  
 XX  
 PS Example; SEQ ID NO 121; 390bp; English.  
 XX  
 CC The present invention relates to the isolation of corn ear-derived  
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022  
 CC and SATMON023. Some of the cdps uniquely identify structural, functional,  
 CC and regulatory genes of corn ear. The polynucleotide sequences are  
 CC useful for detecting cdps in a sample, for producing a corn ear-specific  
 CC profile of gene transcription, for detecting altered gene expression in  
 CC hybrid or hybrid plants, and for screening several molecules for specific  
 CC binding to the polynucleotide. The cdps are useful to identify, isolate,  
 CC or extend identical or related corn-ear nucleic acid sequences from DNA  
 CC libraries, and in nucleic acid amplification or hybridization techniques  
 CC to follow the expression of desirable traits through plant breeding  
 CC programs. Preferably, the cdps are used to identify, evaluate, alter, or  
 CC follow the inheritance of desired characteristics associated with growth,  
 CC and development, disease resistance, environmental adaptability, quality,  
 CC and yield of corn. The cdps are also useful as molecular markers for  
 CC studying inheritance and multigene traits in a plant breeding program.  
 CC The cdps are useful for producing purified corn-ear polypeptides by  
 CC recombinant techniques. They are also useful in diagnostic assays to  
 CC detect or confirm conditions or diseases associated with abnormal levels  
 CC of cdp expression. ABX81661-ABX89140 represent corn ear-derived  
 CC polynucleotides (cdps) of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/psipds/identity.html  
 CC  
 CC Sequence 272 BP; 79 A; 52 C; 68 G; 67 T; 0 U; 6 Other;

[illegible]

CC	binding site of the polynucleotide. The cdpbs are useful to identify, isolate,
CC	or extend identical or related corn-ear nucleic acid sequences from DNA,
CC	libraries, and in nucleic acid amplification or hybridization techniques
CC	to follow the expression of desirable traits through plant breeding
CC	programs. Preferably, the cdpbs are used to identify, evaluate, alter, or
CC	follow the inheritance of desired characteristics associated with growth
CC	and development, disease resistance, environmental adaptability, quality,
CC	and yield of corn. The cdpbs are also useful as molecular markers for
CC	studying inheritance and multigene traits in a plant breeding program.
CC	The cdpbs are useful for producing purified corn-ear polypeptides by
CC	recombinant techniques. They are also useful in diagnostic assays to
CC	detect or confirm conditions or diseases associated with abnormal levels
CC	of cdp expression. ABRX1541-ABX9140 represent corn ear-derived
CC	polynucleotides (cdps) of the invention. Note: The sequence data for this
CC	patent did not form part of the printed specification, but was obtained
CC	in electronic format directly from the USPRO web site at
CC	seqdata.uspro.gov/peipdbIDEntry.html
XX	
XX	Sequence 274 BP; 83 A; 49 C; 73 G; 67 T; 0 U; 2 Other;
XX	
XX	Query Match 4.0%; Score 74.4; DB 7; Length 274;
XX	Best Local Similarity 60.5%; Pred. No. 1.3e-10;
XX	Matches 156; Conservative 0; Mismatches 98; Indels 4; Gaps 2;
OY	1009 TATGAGGCTCTTAAGAGACCGCATTCCTATAGGT--CCACAGGCGCATGTGGAGTAGGT 1067
DB	2 TATGAAACGAGTAAAGACCGCATGCTATAGGTCTCTGGGGACCGTGTGNAAGACGT 61
OY	1068 TCTGATGTTTGAGAGCAGTGGCACTGGCTATTTTGAAGGCGCAAGCCTCCACCGGAGTT 1127
DB	62 GTTAATATTTGAAAGCTCAGCTGTGGGCTCATGGAAGCTGAAAGCTTGCTAAGCACTT 121
OY	1128 AGCTGAGATGGGTTAGATAGAAATTTGCTGG--GGTCAGAAAGCAGATATGTTTTCTGG 1184
DB	122 TGTAAATCAAGGTACAGACAGGAATTCATGGCACTACGAAAGTTGGATTGTGGCTGG 181
OY	1185 AGGTGTTCCGCAACTGTATGCTCTCTTGCACAGACGAAGATTTGGACATATTCAATCA 1244
DB	182 TGGGAAAGGCAACTATATGTGTTCTTACGAAACAAAGAGATATGAGGCAATTTAAACA 241
OY	1245 ACACTCTCAAGGCAAAAC 1262
DB	242 GCATTGCCAGGTAAAGC 259
ID	
ID	RESULT 5
ABLU71922	
ID	ABLU71922 standard; cDNA; 310 BP.
XX	
AC	ABLU71922;
XX	
DT	14-MAY-2002 (first entry)
XX	
DE	Corn tassell-derived polynucleotide (cdps) SEQ ID NO:1296.
XX	
XX	Corn; corn tassell-derived polynucleotide; cdp; hybrid breeding; CDPs;
KW	inheritance; characteristic; growth; development; disease resistance;
KW	environmental adaptability; quality; yield; molecular marker;
KW	multigene trait; plant breeding; corn tassell; gene; ss.
OS	
OS	Zea mays.
XX	
PN	US2001051335-A1.
XX	
PD	13-DEC-2001.
XX	
XX	16-APR-1999; 99US-00294093.
XX	
PR	21-APR-1998; 98US-0082567P.
XX	
PA	(LALG/) LALGUDI R V.
PA	(ITOL/) ITO L Y.
PA	(SHER/) SHERMAN B K.





QY 1278 GTTGAATCATATCAAGAGATGTTGTTAAAGAGCTGAGCAGATCTTGAAGACAAATCA 1337  
DB 1860 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 1919  
QY 1338 GCAGCTGAATCTTTTAAAGACAGCTCTCAAAAAGAAACAGCAGCCAGAGTCTTGA 1397  
DB 1920 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1979  
QY 1398 GGAATCTCTGAAATTTATGAGGAGAGAGCTGGTGAATCTGACAGAGATTAATCGATCGT 1457  
DB 1980 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGATGAGCA 2039  
QY 1458 GAGACAGAGAACTAAGATCAGCAGATGAGCAGCAGCAGAGAGATGAGCAGCAGAG 1517  
DB 2040 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2099  
QY 1518 GTTTTTCATGATTCATTAACAAAGATCCATGAAAGAGACGCAAGAGAGAGATTT 1577  
DB 2100 GCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCA 2159  
QY 1578 CGAGATGTTGACGACGACAGCAGCAGTTCGCAAGGTTGTTGCGCAGCAGCAGCAATTA 1637  
DB 2160 GATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 2219  
QY 1638 TCCCTTAGCAATGACGATTCGCAAGAGAGCTGAGAGAGTGTCAAGCTTCATCGAGTT 1697  
DB 2220 GCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGAGAGAGCA 2279  
QY 1698 TCAAGAGAAAGATGAGAGAGTTTGTGAAGAGAGAGAGATGCTGATTAAGATCAAGA 1757  
DB 2280 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2339  
QY 1758 GAAAGAGATGAGAGCAGTGAAGAGAGCAGCAGAGATTTGATCTGAGAGAAAGA 1817  
DB 2340 GCAGAGATGAGAGCAGCAGCAGATGAGAGAGAGCAGCAGAGATTTAGAGAGCA 2399  
QY 1818 ATTGATGAGCTTTGAGAGCAG 1839  
DB 2400 GCAGCAGAGATGAGAGAGCAG 2421

RESULT 7  
AAF82901  
ID AAF82901 standard; DNA; 3489 BP.

AC AAF82901;

DT 11-SEP-2003 (revised)  
DT 06-AUG-2003 (revised)  
DT 29-JUN-2001 (first entry)

DE Nucleotide sequence of KSHV tethering protein, LANA.

XX Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;  
KM Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;  
XX KSHV; latency-associated nuclear antigen; LANA; ds.

OS Human herpesvirus 8.

XX Key Location/Qualifiers  
FT CDS 1..3489  
FT /\*tag= a

FN WO200125484-A2.

XX 12-APR-2001.

XX 29-SEP-2000; 2000WO-US026908.

XX 01-OCT-1999; 99US-00410399.

XX (UNMI ) UNIV MICHIGAN.

XX  
PI Robertson ES, Colter MA;  
XX  
XX MPI: 2001-281736/29.  
DR P-PSDB; AAB62331.  
XX  
XX A composition for use in gene therapy comprises an expression vector that  
PT includes a nucleic acid sequence encoding a nucleic acid binding protein.  
PS Disclosure; Fig 9A; 60pp; English.  
XX  
XX The invention provides a composition comprising nucleic acid, histone H1  
CC protein and expression vector operably encoding a protein suitable  
CC for tethering the nucleic acid to the histone H1 protein, where the  
CC tethering protein is LANA. The composition is useful in aiding the  
CC retention of the viral DNA in the host cell. The viral vector encodes a  
CC protein suitable for tethering DNA to Histone H1. Methods for screening  
CC for compounds which are agonistic or antagonistic for the tethering of  
CC viral proteins to histone H1 and DNA binding sites are useful for  
CC developing the method of viral transfer. The composition has applications  
CC to gene therapy, including the treatment of multiple sclerosis,  
CC Parkinson's disease, Huntington disease and diabetes. The present  
CC sequence represents the nucleotide sequence of the Kaposi's sarcoma  
CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear  
CC antigen (LANA), which acts as a tethering protein. (Updated on 06-AUG-  
CC 2003 to correct OS field.) (Updated on 11-SEP-2003 to standardise OS  
CC field)  
SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 U; 0 Other;  
Query Match 3.1%; Score 58.8; DB 4; Length 3489;  
Best Local Similarity 43.4%; Pred. No. 1.4e-05;  
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;  
QY 1218 GAAAGAAATTCGACATATTCATCAACCTCTTAAGCAAAAGAGCTGAATTGCA 1277  
DB 1800 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGATGAGCA 1859  
QY 1278 GTTGAATCATATCAAGAGATGTTGTTAAAGAGCTGAGCAGATCTTGAAGACAAATCA 1337  
DB 1860 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGATGAGCA 1919  
QY 1338 GCAGCTGAATCTTTTAAAGACAGCTCTCAAAAAGAAACAGCAGCCAGAGTCTTGA 1397  
DB 1920 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGATGAGCA 1979  
QY 1398 GGAATCTCTGAAATTTATGAGGAGAGCTGGTGAATCTGACAGAGATTAATCGATCGT 1457  
DB 1980 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGATGAGCA 2039  
QY 1458 GAGACAGAGAACTAAGATGAGCAGCAGATGAGCAGCAGAGAGATGAGCAGCAGAG 1517  
DB 2040 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2099  
QY 1518 GTTTTTCATGATTCATTAACAAAGATCCATGAAAGAGACGCAAGAGAGAGATTT 1577  
DB 2100 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2159  
QY 1578 CGAGATGTTGACGACGACAGCAGTTCGCAAGGTTGTTGCGCAGCAGCAGCAATTA 1637  
DB 2160 GATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2219  
QY 1638 TCCCTTAGCAATGACGATTCGCAAGAGAGCTGAGAGAGTGTCAAGCTTCATCGAGTT 1697  
DB 2220 GCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCA 2279  
QY 1698 TCAAGAGAAAGATGAGAGAGTTTGTGAAGAGAGAGATGCTGATTAAGATCAAGA 1757  
DB 2280 GCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGATGAGCA 2339  
QY 1758 GAAAGAGATGAGAGCAGTGAAGAGAGCAGCAGAGATTTGATCTGAGAGAAAGA 1817  
DB 2340 GCAGAGATGAGAGCAGCAGCAGATGAGAGAGAGCAGCAGAGATTTAGAGAGCA 2399

QY	1818	ATTGATGAGGCTTTGGACAG	1839
Db	2400	GGAGCAGGAGTTTGGAGGACG	2421
XX	RESULT 8		
XX	ABA93487		
XX	ABA93487	standard; DNA; 3489 BP.	
AC	ABA93487;		
DT	25-APR-2002	(first entry)	
XX	Kaposi's sarcoma-associated herpesvirus LANA protein encoding DNA.		
XX	Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;		
KW	KSHV terminal repeat; rhadino virus cis acting element; episome;		
KW	primary effusion lymphoma; latency-associated nuclear antigen;		
KW	gene therapy; gene transfer; gene; ds.		
XX	Human herpesvirus 8.		
OS			
XX			
XX	Key	Location/Qualifiers	
FT	CDS	1..3489	
FT		/*tag = a	
FT		/product= "LANA protein"	
FT		/note="latency-associated nuclear antigen"	
XX	US6322792-B1.		
PN	27-NOV-2001.		
PD			
XX	21-APR-1999;	99US-00298568.	
PF			
XX	19-NOV-1998;	98US-0109422P.	
PR			
XX	(KIEF/) KIEFF E D.		
XX			
PA	Kieff ED, Ballestas ME, Kaye KM;		
PI			
XX	WPI; 2002-153769/20.		
XX	P-PSDB; ABB05621.		
DR			
XX	System for episomal retention of plasmids in mammalian cells, useful in		
XX	gene therapy, comprises rhadinoviral LANA and RVCAE sequences.		
XX	Claim 1; Fig 6; 27P; English.		
XX			
XX	The present invention describes a system (A) for maintaining a plasmid as		
XX	an episome in mammalian cells, comprising the rhadinoviral sequence LANA		
CC	(latency-associated nuclear antigen) of 3489 base pairs (see ABA93487,		
CC	(S1) expressed in the cell, and the rhadinoviral sequence RVCAE		
CC	(rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2)		
CC	present in the plasmid. Also described is a method for maintaining a		
CC	closed circular DNA in a cell by expressing (S1) in the cells and having		
CC	(S2) as a cis-acting and maintenance sequence in the DNA. (A) is		
CC	particularly used in gene therapy (or other gene transfer applications)		
CC	that uses mammalian cells in which LANA is expressed. (A) improves		
CC	persistence of gene therapy vectors in cells. The present sequence		
CC	encodes Kaposi's sarcoma-associated herpesvirus (KSHV, also called human		
CC	herpesvirus 8) LANA protein, which is used in the exemplification of the		
CC	present invention		
XX	Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 U; 0 Other;		
QY			
XX	Query Match.	3.11%; Score 58.8; DB 6; Length 3489;	
XX	Best Local Similarity	43.4%; Pred. No. 1.4e-05;	
XX	Matches 270; Conservative	0; Mismatches 352; Indels	0; Gaps
XX			0;
XX	1218	GAAGCAATCTGACATATTCATCACTCTCAAGCGAAACAGGCTGAATTGCA	1277
XX	1800	GCAGCAGATGAGCAGCAGCAGAGATGAGCAGCAGGATGAGCAGCAGGATGACGA	1855

Qy	1278	GTGAATCATATCCAGAGATGGTTGTAAGAAGCTTGAGCGAGATCTCTGAGCAATCA	1333
Db	1860	GCAGCAGGATGAGCAGCAGCAGAGATGACGACGACGAGATGACGACGAGATGACGA	1919
Qy	1338	GCAGCTGAACCTATTTAAGAACAAAGCTCTCAAAACAGAACCAAGCAGCCAGTCTTGA	1397
Db	1920	GCAGCAGGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGAGATGACGA	1979
Qy	1398	GGAATCTCTGAAAATTATGAGCGAAGAGCTGCGTGAATCTGACAGAGATATTCGATCGT	1457
Db	1980	GCAGCAGAGATGACAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGACGA	2039
Qy	1458	GAGACAGAGAACTTAAGATGACGAGATGACAGAACAGAACAGAGAGATGATGACACGACAG	1517
Db	2040	GCAGAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGAGATGACAGCAGCAGAGATGACGA	2099
Qy	1518	GTTTTCATGAGATTCAAATCAAAACAGATCCATGAAAGAAAGACGCAAAAGAGAGAAATTT	1577
Db	2100	GAGCAGCAGAGATGACGAGCAGCAGCAGAGATGACAGCAGCAGCAGCAGATGAGCAGCAGCA	2159
Qy	1578	CGAGATGTTTGAGCAGCAGCAGAACCGTCCAAAGTGTGTTGTCACAGCAGCAGCAAAACATTA	1637
Db	2160	GGATATAGACGACGACGACGAGATGAGCAGCAGCAGCAGAGATGACGACGACGACGAGATGA	2219
Qy	1638	TCCCTCTAGCAATGACGATTTGCCAAAGAGAGCTGAGAAAGTGTCAAGCTTCATCGAGTT	1697
Db	2220	GCAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGATGACAGCAGCAGCAGCAGCAGCAGCAGCA	2279
Qy	1698	TCAAGAGAAAGATGAGCAGAGTTTGTGAAAGAGAGAGAGATGCTGATTAAGATCAAGA	1757
Db	2280	GCAGCAGAGAGCAGCAGCAGCAGAGATGAGAGAGCAGCAGAGATGAGAGATCAGGA	2339
Qy	1758	GAAAGAGATGAAAGACATGAAAGAAAGAGCATCAGCAGAGATATTGATCTCGAGAAAGA	1817
Db	2340	GCAGGAGTTTGAAGGACGACGAGCAGAGATTTAGAGAGCAGCAGCAGAGTTTAAAGAGACA	2399
Qy	1818	ATTGATGAGGCTTTGGAACAG	1839
Db	2400	GAGCAGAGATTAGAGAGCAG	2421
RESULT 9			
AAV73805/c			
ID	AAV73805	standard; DNA; 32207 BP.	
XX	AAV73805;		
AC	17-OCT-2003	(revised)	
DT	25-FEB-1999	(first entry)	
DE	KSHV LTR DNA (nucleotides 105,301-137,507).		
XX	Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;		
XX	dihydrofolate reductase; LTR; long unique region; vaccine; prophylaxis;		
KW	diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;		
KW	glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;		
KW	v-add; G-protein coupled receptor; FGFRAT; ds.		
OS	Human herpesvirus 8.		
XX	US5849564-A.		
PN	15-DEC-1998.		
XX	29-NOV-1996;	96US-00770379.	
XX	29-NOV-1996;	96US-00770379.	
PR	(UYCO ) UNIV COLUMBIA NEW YORK.		
XX	Bohenzky RA, Moore PS, Russo JT, Chang Y, Edelman IS;		
XX			



DR MPI, 1999-069741/06.  
XX Kaposi's sarcoma-associated herpes virus nucleic acid - encodes  
PT dihydrofolate reductase and is useful for treatment, prophylaxis or  
PT diagnosis of Kaposi's sarcoma.  
XX  
PS Disclosure; Col 155-182, 109pp; English.  
XX  
CC This sequence is a fragment of the Kaposi's sarcoma-associated  
CC herpesvirus (KSHV) LVR (long unique region). This fragment contains  
CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67  
CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,  
CC ORF69, K12 which encodes Kaposin, K13, ORF72 which encodes a cyclin D,  
CC ORF73 which encodes immediate early protein (IEP), K14 which encodes  
CC (v-adh), ORF74 which encodes G-protein coupled receptor, ORF75 which  
CC encodes tegument protein/FGFARF, K15. KSHV is a new human Herpesvirus  
CC (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common  
CC form of neoplasm occurring in persons with acquired immune deficiency  
CC syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis,  
CC diagnosis and treatment of a subject with Kaposi's sarcoma and for  
CC detecting expression of a DNA virus associated with Kaposi's sarcoma in a  
CC cell. (Updated on 17-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T; 0 U; 0 Other;  
Query Match 3.1%; Score 58.8; DB 2; Length 32207;  
Best Local Similarity 43.4%; Pred. No. 4.5e-05;  
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;  
QY 1218 GAAGCAAGATCTGGACATATTCATCAACACTCTCAAGCAAAAGGCTGAATTCGA 1277  
DB 20197 GCAGCAGATGAGCAGACAGAGATGAGCAGCAGATGAGCAGCAGAGATGAGCA 20138  
QY 1278 GTTGAATATATCCAGAGATGTTTAAAGAGCTGAGCAGATCTGAGACATCA 1337  
DB 20137 GCAGCAGATGAGCAGACAGAGATGAGCAGCAGATGAGCAGCAGAGATGAGCA 20078  
QY 1338 GCAGCTGAACCTTAAAGAAACAAGCTTCAAAACAGAAACAGCAGCCAGGTCCTTGA 1397  
DB 20077 GCAGCAGATGAGCAGACAGAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCA 20018  
QY 1398 GGAATCTCTGGAATATATGAGCAGAAAGCTGCTAGAACTGCAGAGATATCGATCGT 1457  
DB 20017 GCAGCAGATGAGCAGACAGAGATGAGCAGCAGATGAGCAGCAGAGATGAGCA 19958  
QY 1458 GAGACAGAGAACTAAGATGAGCAGATGAACGAAACAGGAAAGATGATGACACACAG 1517  
DB 19957 GCAGCAGATGAGCAGACAGAGATGAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 19898  
QY 1518 GTTTTTCATGATTCATCAATCAACAGATTCATGAAGAAAGAGCGAAAGAGAGATT 1577  
DB 19897 GCAGCAGCAGAGATGAGCAGAGAGCAGAGATGAGCAGAGCAGCAGAGATGAGCAGCA 19838  
QY 1578 CGAGATGTTGACAGCAGCAGAGAAAGTGCAGAGTTTGGCCAGCAGCAGCAGAACTTAA 1637  
DB 19837 GGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGAGATGA 19778  
QY 1638 TTCCTCTAGCAATGAGATTTGCCGAAAGAGAGCTGAGGAAGTCTCAAGCTTCATCGATT 1697  
DB 19777 GCAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAAGTGAACGAGCAGCAGAGAGCA 19718  
QY 1698 TCAGAGAAAGAGATGAGAGAGTTTGTGAAGAGAGGAGATGCTATAAAGATCAAGA 1757  
DB 19717 GCAGCAGAGAGAGCAGAGAGAGAGATTAAGAGAGCAGAGCAGATTGAGAGATGAGCA 19658  
QY 1758 GAAGAGATGAGAGACATGAAGAGAGCATCAAGAGAGATATTGATCTGAGAGAAAGA 1817  
DB 19657 GCAGAGATTAGAGAGCAGAGAGAGAGATTAGAGAGAGCAGAGAGATTAGAGAGAGCA 19598  
QY 1818 ATTGATGAGGCTTTGGAACAG 1839  
DB 19597 GCAGCAGAGATTAGAGAGAGCAG 19576

RESULT 10  
AAV19941/C  
ID AAV19941 standard; DNA; 137507 BP.  
XX  
AC AAV19941;  
XX  
DT 27-AUG-2003 (revised)  
DT 03-AUG-1998 (first entry)  
XX  
DE KSHV long unique coding region and terminal repeat.  
XX  
XX KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;  
XX interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;  
XX complement-binding protein; glycoprotein; capsid protein IV; infection;  
XX immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;  
XX lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;  
XX HIV immune status; anti-inflammatory agent; therapy; ds.  
XX  
OS Human herpesvirus 8.  
XX  
FH Key Location/Qualifiers  
FT 1142..2794  
FT /tag= a  
FT /product= "complement-binding protein"  
FT 8659..11236  
FT /tag= b  
FT /product= "glycoprotein B"  
FT complement(17261..17875)  
FT /tag= c  
FT /product= "interleukin 6"  
FT complement(21548..21832)  
FT /tag= d  
FT /product= "macrophage inflammatory protein II"  
FT complement(27137..27424)  
FT /tag= e  
FT /product= "interferon regulatory factor 1"  
FT 28661..29741  
FT /tag= f  
FT /product= "protein T1.1"  
FT complement(58976..60175)  
FT /tag= g  
FT /product= "glycoprotein M"  
FT complement(69412..69915)  
FT /tag= h  
FT /product= "glycoprotein L"  
FT complement(88410..88910)  
FT /tag= i  
FT /product= "interferon regulatory factor 2"  
FT 89600..90541  
FT /tag= j  
FT /product= "interferon regulatory factor 3"  
FT 90173..90643  
FT /tag= k  
FT /product= "glycoprotein X"  
FT complement(93636..94127)  
FT /tag= l  
FT /product= "interferon regulatory factor 4"  
FT complement(111931..112443)  
FT /tag= m  
FT /product= "capsid protein IV"  
FT complement(123808..127296)  
FT /tag= n  
FT /product= "immediate early protein"  
XX  
XX WO9804576-A1.  
XX  
XX PD 05-FEB-1998.  
XX  
XX PF 22-JUL-1997; 97WO-US013346.  
XX  
XX PR 25-JUL-1996; 96US-00686243.  
XX  
XX PR 25-JUL-1996; 96US-00686349.

PR 25-JUL-1996; 96US-00686350.  
 PR 05-SEP-1996; 96US-00687253.  
 PR 10-OCT-1996; 96US-00788323.  
 PR 13-NOV-1996; 96US-00747887.  
 PR 13-NOV-1996; 96US-00748640.  
 PR 29-NOV-1996; 96US-00757669.  
 (UYCO) UNIV COLUMBIA NEW YORK.  
 Chang Y, Bohenzky RA, Russo JJ, Edelman IS, Moore PS;  
 WPI; 1996-130615/12.  
 New nucleic acid encoding Kaposi's sarcoma associated herpes virus  
 proteins - useful for, e.g. detecting levels of HHV8 in, and preparation  
 of vaccines for treatment of, HIV patients.

Example 2; Page 135-203; 230pp; English.

This sequence represents the long unique region and terminal repeat of  
 the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known  
 as human herpes virus 8 (HHV8). This sequence selected from: (a) viral  
 invention which encode KSHV polypeptides selected from: (a) viral  
 macropage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);  
 (c) viral IRF 1; (d) complement-binding protein; and (e) immediate early protein  
 encoded by ORF73. Labeled probes for the nucleic acid, proteins encoded  
 by it, and antibodies (Ab) specific for the proteins are useful for  
 detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body  
 fluids or tissue samples. HHV8 infections can be treated with antiseise  
 or triplex forming molecules or agents that bind specifically to the  
 protein. Ab may be used for prophylaxis or treatment of HHV8 infection,  
 while the protein can be used in protective vaccines. Ab may also be used  
 to differentiate between lymphomas, and HHV8 may be implicated in many  
 other lymphoproliferative diseases such as lymphomas, leukaemia,  
 CC splenomegaly and mycosis fungoides. Cells and animals containing the  
 CC nucleic acid are useful for drug screening. HHV8-derived peptidase gene  
 CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene  
 CC can be inhibited with methotrexate. These can also be used to determine  
 CC the immune status of a patient infected with HIV. HHV8 derived protein  
 CC viral MIP II may be used as an anti-inflammatory agent for, e.g.  
 CC treating rheumatoid arthritis. This sequence is stated as containing 81  
 CC open reading frames. (Updated on 27-AUG-2003 to correct OS field.)

Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 U; 0 Other;  
 Query Match 3.1%; Score 58.8; DB 2; Length 137507;  
 Best Local Similarity 43.4%; Pred. No. 9.7e-05;  
 Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

1218 GAGCAAGATCTGCACATTCATCAATCAACACTCTCAAGGCAAAACAGCTGAAATTCGA 1277  
 125497 GCAGCAGATGAGCAGCAGCAGATGACAGCAGATGACAGCAGCAGATGACGA 125438  
 1278 GTTGAATCATCAACAGATGTTGTAAGACCTGACGACGATCTCTGAGACATCA 1337  
 125437 GCAGCAGATGAGCAGCAGCAGATGACAGCAGATGACAGCAGCAGATGACGA 125378  
 1338 GCAGCTGAACCTTTAAGACACAGCTCTCAAAACAGACAGCAGCAGCTGCTTGA 1397  
 125377 GCAGCAGATGAGCAGCAGCAGATGACAGCAGATGACAGCAGCAGATGACGA 125318  
 1398 GGAATCTCGAATTAATGAGCAGAGAGCTGCTGAAACTGAGAGGATTAATCGATCT 1457  
 125317 GCAGCAGATGAGCAGCAGCAGATGACAGCAGATGACAGCAGCAGATGACGA 125258  
 1458 GAGCAGAGAACTTAAGATGACAGATGAACAGAGAGAGATGATGACACAGCAGAG 1517  
 125257 GCAGCAGATGAGCAGCAGCAGATGACAGCAGATGACAGCAGCAGATGACGA 125198  
 1518 GTTTTCATGATTCATCAACAGATTCATGAAGAGAGAGCAGAGAGATTT 1577

125197 GAGCAGCAGATGAGCAGCAGCAGATGACAGCAGATGACAGCAGATGACAGCAGCA 125138  
 1578 CGAGATGTTGACAGCAGCAGCAGCAGATGACAGCAGATGACAGCAGCAGCAGCAGTAA 1637  
 125137 GATGACAGCAGCAGCAGCAGATGACAGCAGCAGCAGATGACAGCAGCAGATGA 125078  
 1638 TCCCTTGAATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1697  
 125077 GCAGCAGCAGCAGATGACAGCAGCAGATGACAGCAGATGACAGCAGCAGCAGCA 125018  
 1698 TCAAGAGAAAGATGAGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1757  
 125017 GCAGCAGCAGCAGCAGCAGCAGATGACAGCAGCAGCAGCAGCAGCAGCAGCAG 124958  
 1758 GAGAGATGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1817  
 124957 GCAGAGATGAG 124898  
 1818 ATTGATGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1839  
 124897 GAGCAGCAGATGAG 124876

RESULT 11

AB199537 standard; cDNA; 1080 BP.

AB199537;

07-MAR-2002 (first entry)

Mouse ischaemic condition related cDNA sequence SEQ ID NO:537.

Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.

Mus musculus.

MO200188188-A2.

22-NOV-2001.

18-MAY-2001; 2001MO-JP004192.

18-MAY-2000; 2000JP-00145977.

(YUNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

Ishikawa K, Arai S, Takahashi Y, Nagata T, Ishii Y;

WPI; 2002-034733/04.

P-PSDB; ABB57221.

Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 expression levels of particular genes defined in the specification or by  
 determining the expression profile of a gene group comprising these  
 genes.

Claim 2; Page 1472-1473; 2690pp; English.

The present invention describes a method for examining ischaemic  
 conditions, comprising measuring the expression levels of particular  
 CC genes (I) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (I). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring the  
 CC expression levels of particular genes (AB199202 to AB199912, encoding the  
 CC protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles for ischaemic condition-improving drugs or  
 CC indicator when screening for ischaemic condition-improving drugs or  
 CC therapeutics for ischaemic diseases. AB199913 and AB199914 represent  
 CC primers for a mouse ischaemic condition related sequence, which are used



XX WPI, 1997-553460/51.  
DR P-PsDB; AAM30763.  
XX Positive regulatory gene of mannose-1-phosphate transfer in yeast  
PT useful for high mannose type neutral saccharide chain production.  
XX  
XX Claim 2; Page 14-17; 23pp; Japanese.  
XX  
XX This sequence represents the gene of the invention, designated MN4,  
CC encoding a protein which positively regulates mannose-1-phosphate  
CC transfer in yeast. The gene is useful for the preparation of human high  
CC mannose type neutral saccharide chain  
SQ Sequence 4000 BP; 1453 A; 750 C; 770 G; 1027 T; 0 U; 0 Other;  
2.9%; Score 55; DB 2; Length 4000;  
Query Match Best Local Similarity 47.9%; Pred. No. 0.00018;  
Matches 191; Conservative 0; Mismatches 205; Indels 3; Gaps 1;  
QY 1441 GAGGATATCGATCGATCGAGACAGAGAACTTAAGATCGACATGAGACAGAAACAGGAGAG 1500  
DB 3539 GAGAAAGAAAGAAAG 3598  
QY 1501 ATGATGACACACGACAGGTTTTCATGATTCATCAATCAACGATCCATGAAAGAGAGC 1560  
DB 3599 AAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 3655  
QY 1561 GCAAAGAGAGAAATTCGAGATGTCAGACAGAGAAAGTCGCAAGGTTGTGCGCAG 1620  
DB 3656 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 3715  
QY 1621 CAGCGCAGAACTTAATCCCTTAGCAATGACATTCGCCAAGAGAGAGAGAGAGAG 1680  
DB 3716 AAGAAAGAAAGAAAG 3775  
QY 1681 TCAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
DB 3776 GAAAGAAATTAAGAAATGAAATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3835  
QY 1741 CTGATTAAGATCAAGAGAAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 1800  
DB 3836 AAGCAGAGAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3895  
QY 1801 TTGATCTGAGAGAAAGATTTGATGAGGCTTTGAAACAG 1839  
DB 3896 GAAAG 3934  
RESULT 14  
AAC40071  
ID AAC40071 standard; DNA; 1140 BP.  
XX AAC40071;  
AC  
XX  
XX 17-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 26940.  
DE  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825P.  
PR  
XX  
XX 05-MAR-1999; 99US-0123180P.  
PR

PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132633P.  
PR 11-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136332P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 16-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0138455P.  
PR 18-JUN-1999; 99US-0138457P.  
PR 18-JUN-1999; 99US-0138457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 21-JUN-1999; 99US-0139899P.  
PR 22-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 23-JUN-1999; 99US-0140655P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 28-JUN-1999; 99US-0140919P.  
PR 29-JUN-1999; 99US-0141287P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.

PR	5-JUL-1999	99US-01440055
PR	16-JUL-1999	99US-0144085P
PR	16-JUL-1999	99US-0144086P
PR	19-JUL-1999	99US-0144325P
PR	19-JUL-1999	99US-0144331P
PR	19-JUL-1999	99US-0144332P
PR	19-JUL-1999	99US-0144333P
PR	19-JUL-1999	99US-0144343P
PR	19-JUL-1999	99US-0144335P
PR	20-JUL-1999	99US-0144352P
PR	20-JUL-1999	99US-0144363P
PR	20-JUL-1999	99US-0144384P
PR	21-JUL-1999	99US-0144414P
PR	21-JUL-1999	99US-0145086P
PR	21-JUL-1999	99US-0145088P
PR	22-JUL-1999	99US-0145085P
PR	22-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145089P
PR	22-JUL-1999	99US-0145098P
PR	22-JUL-1999	99US-0145192P
PR	22-JUL-1999	99US-0145145P
PR	23-JUL-1999	99US-0145218P
PR	23-JUL-1999	99US-0145224P
PR	26-JUL-1999	99US-0145576P
PR	27-JUL-1999	99US-0145513P
PR	27-JUL-1999	99US-0145518P
PR	27-JUL-1999	99US-0145519P
PR	28-JUL-1999	99US-0145581P
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146389P
PR	03-AUG-1999	99US-0147038P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147302P
PR	05-AUG-1999	99US-0147319P
PR	05-AUG-1999	99US-0147360P
PR	06-AUG-1999	99US-0147303P
PR	06-AUG-1999	99US-0147416P
PR	09-AUG-1999	99US-0147493P
PR	09-AUG-1999	99US-0147935P
PR	10-AUG-1999	99US-0148117P
PR	11-AUG-1999	99US-0148119P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149733P
PR	20-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149962P
PR	23-AUG-1999	99US-0149930P
PR	25-AUG-1999	99US-0149930P
PR	25-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151303P
PR	31-AUG-1999	99US-0151338P
PR	01-SEP-1999	99US-0151530P
PR	07-SEP-1999	99US-0152633P
PR	10-SEP-1999	99US-0153070P
PR	13-SEP-1999	99US-0153788P
PR	15-SEP-1999	99US-0154018P
PR	16-SEP-1999	99US-0154039P
PR	20-SEP-1999	99US-0154779P
PR	22-SEP-1999	99US-0155139P
PR	23-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155569P
PR	28-SEP-1999	99US-0156458P
PR	29-SEP-1999	99US-0156596P
PR	04-OCT-1999	99US-0157117P
PR	05-OCT-1999	99US-0157753P

PR	06-Oct-1999	99US-01578655
PR	07-Oct-1999	99US-01580322
PR	08-Oct-1999	99US-01583829
PR	12-Oct-1999	99US-01583696
PR	13-Oct-1999	99US-01592939
PR	13-Oct-1999	99US-01592994
PR	13-Oct-1999	99US-01592955
PR	14-Oct-1999	99US-01593299
PR	14-Oct-1999	99US-01593300
PR	14-Oct-1999	99US-01593319
PR	14-Oct-1999	99US-01596337
PR	14-Oct-1999	99US-01596378
PR	14-Oct-1999	99US-01595848
PR	18-Oct-1999	99US-01607419
PR	21-Oct-1999	99US-01607671
PR	21-Oct-1999	99US-01607672
PR	21-Oct-1999	99US-01607688
PR	21-Oct-1999	99US-01607708
PR	21-Oct-1999	99US-01608144
PR	21-Oct-1999	99US-01608155
PR	21-Oct-1999	99US-01608180
PR	22-Oct-1999	99US-01609802
PR	22-Oct-1999	99US-01609819
PR	22-Oct-1999	99US-01609882
PR	22-Oct-1999	99US-01610404
PR	25-Oct-1999	99US-01614055
PR	25-Oct-1999	99US-01614066
PR	26-Oct-1999	99US-01613597
PR	26-Oct-1999	99US-01613609
PR	26-Oct-1999	99US-01613612
PR	26-Oct-1999	99US-01613200
PR	28-Oct-1999	99US-01619922
PR	28-Oct-1999	99US-01619933
PR	29-Oct-1999	99US-01621422

Query Match	2.8%	Score 53.2;	DB 3;	Length 1140;
Best Local Similarity	45.2%;	Pred. No. 0.0003;		
Matches 196; Conservative	0;	Mismatches 238;	Indels 0;	Gaps 0;

QY	1381	AAGCCCAAGTGCCTTGAGAAATCTCTGGAAATTATGAGCCAGAAAGCTCCGTAAGATCGCA	14410
Db	473	CACAGGATGATGGTGTGACGATGCTCACAAAAACATAGTGGTAAAGGATTTATGATGATGGA	532
QY	1441	GAGGATATATCGATCGTGCAGACAGAAACTTAAGTGCAGCATGAACAGAGCGGAAGAG	1500
Db	533	GATGAGAAAGTAAAGAAAGAGAAAGAGAAAGAAAGATTAAAGAAAGATGTTAT	592
QY	1501	ATGATGCACACGACAGGTTTTTCATGATTCATTAACAGATCCATGAAGAAGAGAC	1560
Db	593	AACTCTGAAGTGTATGATTTAAGAAAGAAAGAAATAGCGTCAAGAGACATCAT	652
QY	1561	GCAAGGAGGAGATTTTCAGATGTTGCAGCAGAGAAAGTGCAGAGTTGTTGGCCAG	1620
Db	653	GATGATGATGATTTTATGATGAGAAAGAAAGAAAGAAAGCTTTAATGATGATGATGAG	712
QY	1621	CAGCAGCAGAACATTAAATCCCTCTAGCAATGACATTTGCCGAAAGAGAGCTGAGAAAGTG	1680
Db	713	AAGAAAGAAAGAAAGCATTAATTAATGATGATGATGATGATGAGAAAGAAAGACATTAACAT	772
QY	1681	TCAAGCTTCATCCGAGTTTCAAGAGAAAGATGAGAGAGTTTGTGGAAGAGAGGAATG	1740
Db	773	GATGATGATGATGAGAAAGAAAGAAAGAAAGAGATATCATGATGTGAGGATTAAGAAAGAG	832
QY	1741	CTGATTAAGAGATCAAGAGAAAGAGATGAGAGACATGTGAAGAAAGAGCATCAAGAGAGATA	1800
Db	833	AAGAAAGCATTAATTAATGATGATGATGAGAAAGAAAGAAAGAGATCAATCTGATGAT	892
QY	1801	TTTGATCTGGAGAA	1814
Db	893	GATGATGAGAGAA	906

RESULT 15  
AAx83006/c  
ID AAX83006 standard; DNA; 16442 BP.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 14:52:55 ; Search time 7073.65 Seconds  
(without alignments)  
11507.244 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878

Sequence: 1 atgagctcctagcgcctgctcc.....ctccacatgaagatcgtga 1878

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_ey:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1878	100.0	1878	6	AX078761	AX078761 Sequence
2	1878	100.0	1909	8	BT004380	BT004380 Arabidops
3	1878	100.0	2162	8	BT002944	BT002944 Arabidops
4	1462	77.8	3275	6	AF239719	AF239719 Arabidops
5	1462	77.8	3275	6	AX078760	AX078760 Sequence
6	1462	77.8	81365	8	AB025633	AB025633 Arabidops
7	437.8	23.3	650	8	ATH528171	ATH528171 Arabidops
8	427.4	22.8	2181	8	AK064995	AK064995 Oryza sat
9	372	19.8	2570	8	AK100699	AK100699 Oryza sat
10	368	19.6	1556	8	AK064217	AK064217 Oryza sat
11	291.6	15.5	1564	8	AF542974	AF542974 Trilicium
12	288.8	15.4	1214	8	AK061975	AK061975 Oryza sat
13	192	10.2	513	8	AF469493	AF469493 Trilicium
14	169.4	9.0	91053	8	CNS086CAO	AL831811 Oryza sat
15	169.4	9.0	112231	8	CNS077QI	AL713901 Oryza sat
16	130.4	6.9	94911	8	ATT14E10	AL138656 Arabidops
17	88.4	4.7	179685	10	AC122056	AC122056 Mus muscu
18	87.8	4.7	7218	6	I66494	I66494 Sequence 14
19	86.6	4.6	204635	10	AL928594	AL928594 Mouse DNA
20	84	4.5	282861	2	AC111701	AC111701 Rattus no
21	83.4	4.4	227272	2	AC103478	AC103478 Rattus no
22	83.2	4.4	279587	2	AC112408	AC112408 Rattus no
23	82.4	4.3	157393	10	AC124199	AC124199 Mus muscu
24	80.2	4.3	246672	2	AC105804	AC105804 Rattus no
25	79.4	4.2	124244	10	AL929026	AL929026 Mouse DNA
26	79.4	4.2	188782	10	AL713870	AL713870 Mouse DNA
27	79	4.2	195690	10	AC098726	AC098726 Mus muscu
28	78.6	4.2	114848	9	AC018664	AC018664 Homo sapi
29	78.6	4.2	178635	9	AC008080	AC008080 Homo sapi
30	78.6	4.2	194006	10	AL732441	AL732441 Mouse DNA
31	78	4.2	233713	2	AC133034	AC133034 Rattus no
32	78	4.2	237549	2	AC112451	AC112451 Rattus no
33	77.8	4.1	201074	2	AC102097	AC102097 Mus muscu
34	77.8	4.1	211947	10	AC107828	AC107828 Mus muscu
35	77.6	4.1	188587	10	AC124464	AC124464 Mus muscu
36	77.6	4.1	205037	10	AC126266	AC126266 Mus muscu
37	77.4	4.1	159256	10	AC127264	AC127264 Mus muscu
38	77.4	4.1	178927	2	AC118897	AC118897 Rattus no
39	77.4	4.1	217511	10	AC125357	AC125357 Mus muscu
40	77.2	4.1	56052	8	AP006363	AP006363 Lotus cor
41	77	4.1	215910	10	AC127335	AC127335 Mus muscu
42	76.8	4.1	195987	10	AC122537	AC122537 Mus muscu
43	76.8	4.1	197599	2	AC119769	AC119769 Rattus no
44	76.8	4.1	226601	10	AC122863	AC122863 Mus muscu
45	76.6	4.1	248718	10	AC117629	AC117629 Mus muscu

## ALIGNMENTS

RESULT 1  
LOCUS AX078761 1878 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 2 from Patent WO0105951.  
ACCESSION AX078761  
VERSION AX078761.1 GI:13158380  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS Beclin,C., Elmayan,T. and Vaucheret,H.  
TITLE Novel sgsl plant gene and use thereof

## JOURNAL

Patent: WO 0105951-A 2 25-JAN-2001;  
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE  
 AGRONOMIQUE (FR)

## FEATURES

## Source

Location/Qualifiers

1..1878

/organism="Arabidopsis thaliana"

/mol\_type="unassigned DNA"

/db\_xref="taxon:3702"

## CDS

1..1878

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC32419.1"

/db\_xref="GI:13158381"

/db\_xref="EMBL:CA032419"

translation="MSRAQPMSEKENVVGGYRPEVQLVQGLAGTRLIASSQDDGEG  
 EISKNNKPKPNTSGKTWVSONSNPRAVSGQOQSGNSVNGNNGVSGNNGRG  
 IOANTSGRGALSRVYDNNFVAAPVSRPLEGGMNMOARGSAHTVQFPEVED  
 VNASRENDSDALDDSDDDLASDSDSDVSKSHSGRKNKKFKFGLSDLSIEQ  
 INEPRQWHPACQNGPQALIDYINLHPLAARTKARVYLHRELALEKQDLEMY  
 ASVIPCGLYQMGKIGDEKDYELVPEPVYINNTLRDKDNDKMLQMOLELY  
 DYKALRAHSHYQPGHGMSTVMESSATGYLEARLHRELALEKQDLEMY  
 FSGVROLVGLFATKODLDIFNOSHQGRLEKSYQEMVWKELRQISDNOQNT  
 PNKLSKONKHAQVLEBSLEIMSEKLRARIEDNRIVRORTKQHEONBEMADHFRF  
 MDSIKQIHRBRPAKEENPEMLQOORAVVGGQONINPSNDCKRKAEEVSEFIEF  
 OKEMEPEVEEREMLIKQEKEMEMKRNHEIFDLKEPDEALQMLMYKHGLNED  
 D"

## ORIGIN

Query Match 100.0%; Score 1878; DB 6; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTTCTAGAGGCTGTCATATGTTAAAGAAAAGACGTTGAGGTTATAGGCT 60  
 1 ATGAGTTCTAGAGGCTGTCATATGTTAAAGAAAAGACGTTGAGGTTATAGGCT 60  
 61 GAGGTTGAACAGTTGGTTCAAGTTGGGACGAGACGAGCTTCTTCAACAAGTAT 120  
 61 GAGGTTGAACAGTTGGTTCAAGTTGGGACGAGACGAGCTTCTTCAACAAGTAT 120  
 121 GAGAGAGAGTGGAGGTCATTTCCAGAGAACAGAACAAACAGAGAAACCTTTGGA 180  
 121 GAGAGAGAGTGGAGGTCATTTCCAGAGAACAGAACAAACAGAGAAACCTTTGGA 180  
 181 AAAAATTGGTTCTCAAAATTCCTTCAGAGCTTGGGTTGTCAGAGCAAGG 240  
 181 AAAAATTGGTTCTCAAAATTCCTTCAGAGCTTGGGTTGTCAGAGCAAGG 240  
 241 AGAGGTAGCAACGATCTGGGAGAGAAACATGATCCGGAGAGGTTAAGCAATG 300  
 241 AGAGGTAGCAACGATCTGGGAGAGAAACATGATCCGGAGAGGTTAAGCAATG 300  
 301 CGGGGCAATTCAGATCATATCTGTCGGGAGACGAGCTTGGACAGAAAGTATAC 360  
 301 CGGGGCAATTCAGATCATATCTGTCGGGAGACGAGCTTGGACAGAAAGTATAC 360  
 361 AACTTTGGTGGACCCCACTGATCTCGCCCTCTTGGAAAGAGAGTGAATTGGCAG 420  
 361 AACTTTGGTGGACCCCACTGATCTCGCCCTCTTGGAAAGAGAGTGAATTGGCAG 420  
 421 GCAAGAGAGAGTCTGCTCAGACACAGCTGTGACAGAGTTCTCTGACGTGAGAGAT 480  
 421 GCAAGAGAGAGTCTGCTCAGACACAGCTGTGACAGAGTTCTCTGACGTGAGAGAT 480  
 481 GTGATAATGCTTCTGAGAGAGAGATGATCCAGTCTTTGATGATGATGATGATG 540  
 481 GTGATAATGCTTCTGAGAGAGAGATGATCCAGTCTTTGATGATGATGATGATG 540  
 541 CTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 541 CTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGTGATGATGATGATGATGATGAT 660

601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGTGATGATGATGATGATGATGAT 660  
 661 GAACCAAGAGGAGATGGCAATGTCAGCTTGTGACAAAGGACCTGGTGCATGATGG 720  
 661 GAACCAAGAGGAGATGGCAATGTCAGCTTGTGACAAAGGACCTGGTGCATGATGG 720  
 721 TATACCTGACCTCTCACTAGCTGACAGAGCAAAAGAGACTAGGCGATTAAGCTC 780  
 721 TATACCTGACCTCTCACTAGCTGACAGAGCAAAAGAGACTAGGCGATTAAGCTC 780  
 781 CATAGAAATTTGCTGAGATTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGAT 840  
 781 CATAGAAATTTGCTGAGATTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGAT 840  
 841 CTTTGTGTGATGATTTATGAGGAGTGGAAAGGTTTGGGTGAGAGATGAAAAGATTGAA 900  
 841 CTTTGTGTGATGATTTATGAGGAGTGGAAAGGTTTGGGTGAGAGATGAAAAGATTGAA 900  
 901 ATTGCTGAGCTCCATGATGTCATCATGATGATGATGATGATGATGATGATGATG 960  
 901 ATTGCTGAGCTCCATGATGTCATCATGATGATGATGATGATGATGATGATGATG 960  
 961 AAGTGGCTGGCATGGGCAACCAAGAGCTGTGAAATCTTGCACAAAGTGAAGGCTCT 1020  
 961 AAGTGGCTGGCATGGGCAACCAAGAGCTGTGAAATCTTGCACAAAGTGAAGGCTCT 1020  
 1021 AGAGACGCGCATTCCTATGTCACAGAGGCTGAGTGTGATGATGATGATGATGATG 1080  
 1021 AGAGACGCGCATTCCTATGTCACAGAGGCTGAGTGTGATGATGATGATGATGATG 1080  
 1081 AGCAGTGCCTGCTGCTATTTGAGAGCGCAAGCTCTCAACGGAGTTAGCTGATGATG 1140  
 1081 AGCAGTGCCTGCTGCTATTTGAGAGCGCAAGCTCTCAACGGAGTTAGCTGATGATG 1140  
 1141 TTAGATTAATTTGCTGAGGTCAGAAAGGAGTATGTTTCTGAGAGGTTGCGCAATG 1200  
 1141 TTAGATTAATTTGCTGAGGTCAGAAAGGAGTATGTTTCTGAGAGGTTGCGCAATG 1200  
 1201 TATGCTCTCTTGCACAGAACAGATCTGACATATTCATCAACTCTCAAGGCAAA 1260  
 1201 TATGCTCTCTTGCACAGAACAGATCTGACATATTCATCAACTCTCAAGGCAAA 1260  
 1261 ACAAGGCTGAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320  
 1261 ACAAGGCTGAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320  
 1321 ATCTCTGAGCAATTCAGAGCTGAACTTAAAGAAAGAGCTCTCAAAACAGAAAG 1380  
 1321 ATCTCTGAGCAATTCAGAGCTGAACTTAAAGAAAGAGCTCTCAAAACAGAAAG 1380  
 1381 CAGCGCAAGGCTCTGAGGATCTCTGAAATTTAGCGAGAGCTGCGTGAACCTGCA 1440  
 1381 CAGCGCAAGGCTCTGAGGATCTCTGAAATTTAGCGAGAGCTGCGTGAACCTGCA 1440  
 1441 GAGATTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
 1441 GAGATTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
 1501 ATGATGACACAGACAGGTTTTCATGATGATGATGATGATGATGATGATGATGATG 1560  
 1501 ATGATGACACAGACAGGTTTTCATGATGATGATGATGATGATGATGATGATGATG 1560  
 1561 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAACTGTCAGAGTTTGGCCAG 1620  
 1561 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAACTGTCAGAGTTTGGCCAG 1620  
 1621 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAACTGTCAGAGTTTGGCCAG 1680  
 1621 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAACTGTCAGAGTTTGGCCAG 1680  
 1681 TCAAGCTTCATGAGATTTCAAGAGAAAGATGAGAGATTTTGTGAGAGAGAGAGATG 1740



Db 1681 TCAAGCTTCATGAGCTTTCAGAGAAAGATGAGAGAGCTTTGTGAGAGAGAGAGATG 1740

Qy 1741 CTGATTAAGATCAAGAGAGATGAGAGATGAGATGAGATGAGATGAGATGAGATG 1800

Db 1741 CTGATTAAGATCAAGAGAGATGAGAGATGAGATGAGATGAGATGAGATGAGATG 1800

Qy 1801 TTTGATCTGAGAGAAAGATTTGATGAGCTTTGAGAACAGCTGATGACAGATGAGCTT 1860

Db 1801 TTTGATCTGAGAGAAAGATTTGATGAGCTTTGAGAACAGCTGATGACAGATGAGCTT 1860

Qy 1861 CACATGAGAGATGATGA 1878

Db 1861 CACATGAGAGATGATGA 1878

RESULT 2

LOCUS BT004380 1909 bp mRNA linear PLN 14-FEB-2003

DEFINITION Arabidopsis thaliana clone U20243 unknown protein (At5g23570) mRNA, complete cds.

ACCESSION BT004380

VERSION BT004380.1 GI:28393932

KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

1 (bases 1 to 1909)

REFERENCE 1 (bases 1 to 1909)

AUTHORS Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Huan, V.W., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Tortum, M., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

TITLE Arabidopsis Open Reading Frame (ORF) Clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1909)

AUTHORS Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Huan, V.W., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Tortum, M., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

TITLE Direct Submission

JOURNAL Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

COMMENT The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

FEATURES

source location/Qualifiers

1..1909

/organism="Arabidopsis thaliana"

/mol\_type="mRNA"

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.

Query Match 100.0%; Score 1878; DB 8; Length 1909;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

3'UTR

1 ATGAGTTTACGGCTGTGCTCAATGCTTAAAGAAAGAAAGCTTACGGCTTATAGCCT 60

1 ATGAGTTTACGGCTGTGCTCAATGCTTAAAGAAAGAAAGCTTACGGCTTATAGCCT 60

61 GAGGTTGAACGTTGCTTCAAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

61 GAGGTTGAACGTTGCTTCAAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

Db 61 GAGGTTGAACGTTGCTTCAAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

Qy 121 GGAG 180

121 GGAG 180

Db 121 GGAG 180

Qy 181 AAAAGCTTGGCTTTCAGAAATTCGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 240

181 AAAAGCTTGGCTTTCAGAAATTCGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 240

Db 181 AAAAGCTTGGCTTTCAGAAATTCGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 240

Qy 241 AGAGTTACCAAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

241 AGAGTTACCAAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

Db 241 AGAGTTACCAAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

Qy 301 CGGGGCTTCAAGCTTTCAGAAATTCGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 360

301 CGGGGCTTCAAGCTTTCAGAAATTCGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 360

Db 301 CGGGGCTTCAAGCTTTCAGAAATTCGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 360

Qy 361 AAAGTTGTGGACCCCACTGTATCTGCGCTCTTTGAGAGAGAGAGAGAGAGAGAGAG 420

361 AAAGTTGTGGACCCCACTGTATCTGCGCTCTTTGAGAGAGAGAGAGAGAGAGAGAG 420

Db 361 AAAGTTGTGGACCCCACTGTATCTGCGCTCTTTGAGAGAGAGAGAGAGAGAGAGAG 420

Qy 421 GCAAG 480

421 GCAAG 480

Db 421 GCAAG 480

Qy 481 GTGATTAATGCTTTCGAG 540

481 GTGATTAATGCTTTCGAG 540

Db 481 GTGATTAATGCTTTCGAG 540

Qy 541 CTGCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

541 CTGCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

Db 541 CTGCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 601 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGGATPAGCTTGTGCATCGACGATTAAT 660  
 DB 601 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGGATPAGCTTGTGCATCGACGATTAAT 660  
 QY 661 GAACCAACAGAGGCAATGGCATTGTCCAGCTTGTGAGAACGAGCTGTGTCATCGATTGG 720  
 DB 661 GAACCAACAGAGGCAATGGCATTGTCCAGCTTGTGAGAACGAGCTGTGTCATCGATTGG 720  
 QY 721 TATTAACCTGCAACCTCTACTAGCTCATGCGAGCAAAAAGAGCTAGGCGAGTTAAGCTC 780  
 DB 721 TATTAACCTGCAACCTCTACTAGCTCATGCGAGCAAAAAGAGCTAGGCGAGTTAAGCTC 780  
 QY 781 CATAGAGAAATGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGTCAAT 840  
 DB 781 CATAGAGAAATGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGTCAAT 840  
 QY 841 CCTTGTGTGAGATTTATGAGGCACTGAGAGGCTTGGGAGAGATGAAAAGATTTATGAA 900  
 DB 841 CCTTGTGTGAGATTTATGAGGCACTGAGAGGCTTGGGAGAGATGAAAAGATTTATGAA 900  
 QY 901 ATTGTCTGGCTCCAAATGATCATCATGATTAATGATGATGATGATGATGATGATGAT 960  
 DB 901 ATTGTCTGGCTCCAAATGATCATCATGATTAATGATGATGATGATGATGATGATGAT 960  
 QY 961 AAGTGGCTCGGATGAGGCAACCAAGAGCTGTGAAATPACTTGAACAAGATGAGGCTCT 1020  
 DB 961 AAGTGGCTCGGATGAGGCAACCAAGAGCTGTGAAATPACTTGAACAAGATGAGGCTCT 1020  
 QY 1021 AGAGAGGCGCATTCCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 DB 1021 AGAGAGGCGCATTCCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 AGAGAGGCGCATTCCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 DB 1081 AGAGAGGCGCATTCCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 QY 1141 TTAAGTGAATTCCTGAGGAGTGAAGAGGAGTGTGTTTGTGAGAGTGTGTTTGTGAG 1200  
 DB 1141 TTAAGTGAATTCCTGAGGAGTGAAGAGGAGTGTGTTTGTGAGAGTGTGTTTGTGAG 1200  
 QY 1201 TATGAGTGTGTTTGTGAGAGTGAAGAGGAGTGTGTTTGTGAGAGTGTGTTTGTGAG 1260  
 DB 1201 TATGAGTGTGTTTGTGAGAGTGAAGAGGAGTGTGTTTGTGAGAGTGTGTTTGTGAG 1260  
 QY 1261 ACAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 DB 1261 ACAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 QY 1321 ATCTCTGAGGATCAATGAGAGTGAAGAGGAGTGTGTTTGTGAGAGTGTGTTTGTGAG 1380  
 DB 1321 ATCTCTGAGGATCAATGAGAGTGAAGAGGAGTGTGTTTGTGAGAGTGTGTTTGTGAG 1380  
 QY 1381 CAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 DB 1381 CAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 QY 1441 GAGGATTAATCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
 DB 1441 GAGGATTAATCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
 QY 1501 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
 DB 1501 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
 QY 1561 GCAAGAGAGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
 DB 1561 GCAAGAGAGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
 QY 1621 CAGCAGAGAGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
 DB 1621 CAGCAGAGAGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680

QY 1681 TCAAGCTTCATGAGTTTCAAGAGAGAGATGAGAGGATTTGTGAGAGAGAGAGATG 1740  
 DB 1681 TCAAGCTTCATGAGTTTCAAGAGAGAGATGAGAGGATTTGTGAGAGAGAGAGATG 1740  
 QY 1741 CTGATTAAGATCAAGAGAGAGAGATGAGAGAGATGAGAGAGAGATGAGAGAGATG 1800  
 DB 1741 CTGATTAAGATCAAGAGAGAGAGATGAGAGAGATGAGAGAGAGATGAGAGAGATG 1800  
 QY 1801 TTTGATCTGAGAGAGATTTGATGAGAGGCTTTGAGAGAGATGATGATGATGATGATGAT 1860  
 DB 1801 TTTGATCTGAGAGAGATTTGATGAGAGGCTTTGAGAGAGATGATGATGATGATGATGAT 1860  
 QY 1861 CACAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
 DB 1861 CACAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
 RESULT 3  
 BT002944 2162 bp mRNA linear PLN 15-JAN-2003  
 LOCUS Arabidopsis thaliana clone RAFL14-93-K05 (R20243) unknown protein  
 DEFINITION (AC5623570) mRNA, complete cds.  
 ACCESSION BT002944  
 VERSION BT002944.1 GI:27754622  
 KEYWORDS FL1 CDNA.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eunotsids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 2162)  
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,  
 Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Arabidopsis Full Length cDNA Clones  
 Unpublished  
 2 (bases 1 to 2162)  
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,  
 Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Direct Submission  
 Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan  
 Street, Albany, CA 94710, USA (GSC) members carried out the  
 RIKEN Genomic Sciences Center  
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN  
 Arabidopsis Full-length cDNA'); Seki,M., Narusaka,M., Ishida,J.,  
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
 Hayashizaki,Y. and Shinozaki,K.  
 The Salk, Stanford, PGEC (SPP) Consortium members carried out the  
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M.,  
 Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Wong,C., Wu,H.C.,  
 Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G.,  
 Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,  
 Palm,C.J., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Yamada,K. (SPP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to  
 this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SPP/PGEC)  
 contributed equally to this work as PIs.  
 Annotation based on July 2002 version of the Arabidopsis genome  
 submitted to Genbank.  
 Location/Qualifiers  
 1..2162  
 FEATURES  
 source

```
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAF14-93-K05 (R20243)"
/ecotype="Columbia"
/notes="This clone is in a modified pBluescript vector2
(lambda PS) as a BamHI/XhoI insert."
1..2162
/gene="At5g23570"
1..91
/gene="At5g23570"
92..1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AA022757.1"
/db_xref="GI:27754623"
/translation="MSRAGPMSKEKNYVGYPREVDLVGLAETRLASDQDQGEW
EVIKKNKPKGNTSGKTMVSONSNPNPAMGQOQGRSNTSGRNTVGRNGVGRG
IOANISGRBALSRKYDNNFYAPPVSRPLEGKNNMQARGSNHTVAVQEPDVED
VNASSEBNDSDALDDDDLASDDYDSVSKSHSRKQNKWFKFVGLDLSIEQ
INBPQRMHCPAQONGPGAIDMYNLPLLAHARTGAARVKLHRELAIVLEKDLQMRG
ASVIPCGBIYQGMKGLGDEKDYELVMPMVIMNTRLDKDNDKMLGAGNOLIEYF
DKYEALRAHSGYQGHGMSYLMFESSATGLAEHLRELAEGDLRIAMGQRSM
PSGVRQLYGLATKODLDIPNOSHOGTRLEKLSYQEWVYKELROI SEDNOLNY
FKNTLSKONKHAHYBESLEINSEKLRITAEENRIVRQRTKQHQENBENDHRRP
WDSIKQIHERDKAENFEMLOQDERAVGQOONINPSSNDCKRAEVSSTIEF
QEKEMEEFVEREMLIKQEKEMEDKRHEEIFDLKEFDEALEOLMYKRLHNEED
D"
1970..2162
/gene="At5g23570"
2146
/gene="At5g23570"
/notes="not present in genomic sequence"

ORIGIN
Query Match 100.0%; Score 1878; DB #: Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGGCTGTGTCATGTCTAAGAAAAGAAAGTTCAAGGTTGTTATAGGCT 60
DB 92 ATGAGTTCTAGGGCTGTGTCATGTCTAAGAAAAGAAAGTTCAAGGTTGTTATAGGCT 151
QY 61 GAGGTTGAACAGTTGTTCAAGGTTTGCAAGGAGCAAGACTGGCTTTTCAAGATGAT 120
DB 152 GAGGTTGAACAGTTGTTCAAGGTTTGCAAGGAGCAAGACTGGCTTTTCAAGATGAT 211
QY 121 GAGGAGAGTGGAGAGTCAATTTCCAAGAAACAAGAAACAACCAAGAAACCTCTGGA 180
DB 212 GAGGAGAGTGGAGAGTCAATTTCCAAGAAACAAGAAACAACCAAGAAACCTCTGGA 271
QY 181 AAAAATTGGGTTTCTCAAGATTCGATCTCTTGAAGCTTGGGTTGTCAGACAGAGG 240
DB 272 AAAAATTGGGTTTCTCAAGATTCGATCTCTTGAAGCTTGGGTTGTCAGACAGAGG 331
QY 241 AGAGTAGCAACGATCTGGAGAGAGAAACAATGATCCGGAGAGGTTAACCGCAATGCT 300
DB 332 AGAGTAGCAACGATCTGGAGAGAGAAACAATGATCCGGAGAGGTTAACCGCAATGCT 391
QY 301 CGGGGCAATTCAGATCAATATCTGGTGGGGGACAGAGGTTGACGAAAGATATGATAC 360
DB 392 CGGGGCAATTCAGATCAATATCTGGTGGGGGACAGAGGTTGACGAAAGATATGATAC 451
QY 361 AACTTTGTGACACCCCACTGTATCTGCCCTCTTTTGAAGAGAGATGGAATTTGGAG 420
DB 452 AACTTTGTGACACCCCACTGTATCTGCCCTCTTTTGAAGAGAGATGGAATTTGGAG 511
QY 421 GCAGAGAGAGGTTGCTCAGACACAGCTGTGACAGAGTTTCTTGAAGTGAAGATAT 480
DB 512 GCAGAGAGAGGTTGCTCAGACACAGCTGTGACAGAGTTTCTTGAAGTGAAGATAT 571
```

```
QY 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGATGATGATGACGAC 540
DB 572 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGATGATGATGACGAC 631
QY 541 CTGGCAAGTGAATGATTAATGACTCGGATGATCAAAAGAGCATGATCAAGAAAGCAG 600
DB 632 CTGGCAAGTGAATGATTAATGACTCGGATGATCAAAAGAGCATGATCAAGAAAGCAG 691
QY 601 AATAAGTGTCAAAAGATTTCTTTGGCAGCTTGATCTGTGATGAGAGATTAAT 660
DB 692 AATAAGTGTCAAAAGATTTCTTTGGCAGCTTGATCTGTGATGAGAGATTAAT 751
QY 661 GAACCAAGAGGAGTGGCATTTGTCAGCTTGTCAAGAGGACCTGTGCTCATTTGG 720
DB 752 GAACCAAGAGGAGTGGCATTTGTCAGCTTGTGTCAGAGGACCTGTGCTCATTTGG 811
QY 721 TATAAAGTGAACCTCTAAGCTGATGCGAGGACAAAGAGGCTAAGGAGTTAAGCTC 780
DB 812 TATAAAGTGAACCTCTAAGCTGATGCGAGGACAAAGAGGCTAAGGAGTTAAGCTC 871
QY 781 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGTCA 840
DB 872 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGTCA 931
QY 841 CTTGTGGAGAGATTTATGGGCAAGTGAAGGTTTGGGAGAGATGAAGAAAGATTAATGA 900
DB 932 CTTGTGGAGAGATTTATGGGCAAGTGAAGGTTTGGGAGAGATGAAGAAAGATTAATGA 991
QY 901 ATTGTCTGAGCTCCAGATGATCATCATGATTAAGATTAAGAGCATTAAGAT 960
DB 992 ATTGTCTGAGCTCCAGATGATCATCATGATTAAGATTAAGAGCATTAAGAT 1051
QY 961 AAGTGGCTCGGATGAGGCAACCAAGAGCTGTGGAATCTTGACAAAGATTAAGGCTCT 1020
DB 1052 AAGTGGCTCGGATGAGGCAACCAAGAGCTGTGGAATCTTGACAAAGATTAAGGCTCT 1111
QY 1021 AGAGCAAGCCATCTCTATGATGTCACAGGAGCCATGCTGGAGATGAGTCTGATTTTGA 1080
DB 1112 AGAGCAAGCCATCTCTATGATGTCACAGGAGCCATGCTGGAGATGAGTCTGATTTTGA 1171
QY 1081 AGCAGTGCACATGAGCTATTTGAGAGGCCAGACGCTTCCACCGGAGTTAGCTGAGATGG 1140
DB 1172 AGCAGTGCACATGAGCTATTTGAGAGGCCAGACGCTTCCACCGGAGTTAGCTGAGATGG 1231
QY 1141 TTGATTAAGATTTGCTGGGTTGCAAGAGCGCATGATTTTCTGAGAGTGTGCGCAATG 1200
DB 1232 TTGATTAAGATTTGCTGGGTTGCAAGAGCGCATGATTTTCTGAGAGTGTGCGCAATG 1291
QY 1201 TATGCTTCTTGGCAAGAGAGATCTGGACATTAATCAATCAACATCTCAAGGCAAA 1260
DB 1292 TATGCTTCTTGGCAAGAGAGATCTGGACATTAATCAATCAACATCTCAAGGCAAA 1351
QY 1261 ACAAGCTGAATTCGATGTAATCATACAGAGATGTTGTAAGAGCTGAGCGAG 1320
DB 1352 ACAAGCTGAATTCGATGTAATCATACAGAGATGTTGTAAGAGCTGAGCGAG 1411
QY 1321 ATCTTGAAGCAATCAGACGCTGAATCTTTAAGAACAAAGCTTCAAAACAGAACAG 1380
DB 1412 ATCTTGAAGCAATCAGACGCTGAATCTTTAAGAACAAAGCTTCAAAACAGAACAG 1471
QY 1381 CAAGCAGAGTGTGTAAGAAATCTGTGAATTTAAGACGAGAGCTGCTGAAGCTGCA 1440
DB 1472 CAAGCAGAGTGTGTAAGAAATCTGTGAATTTAAGACGAGAGCTGCTGAAGCTGCA 1531
QY 1441 GAGGATTAATCGGATCTGTGAGACAGAGAACTAAGATGACAGATGAACAGAGAGAG 1500
DB 1532 GAGGATTAATCGGATCTGTGAGACAGAGAACTAAGATGACAGATGAACAGAGAGAG 1591
QY 1501 ATGATGACACGACAGGTTTTTCAATGATTAATCAACAGATTCATGAAGAAAGAGAC 1560
DB 1592 ATGATGACACGACAGGTTTTTCAATGATTAATCAACAGATTCATGAAGAAAGAGAC 1651
```

QY 1561 GCAAGAGAGAAATTTGAGATTTTGCACAGAGAAAGCTCCAGGTTGTGGCCAG 1620  
 DB 1652 GCAAGAGAGAAATTTGAGATTTTGCACAGAGAAAGCTCCAGGTTGTGGCCAG 1711  
 QY 1621 CAGCGAGAGAAATTTTCCCTTACGATGACGATTTGCCGAAAGAGCTGAGGAAATG 1680  
 DB 1712 CAGCGAGAGAAATTTTCCCTTACGATGACGATTTGCCGAAAGAGCTGAGGAAATG 1771  
 QY 1681 TCAGCTTCATGAGATTTTCAAGAGAAAGATGAGGAGTTGTGAAAGAGAGGAGATG 1740  
 DB 1772 TCAGCTTCATGAGATTTTCAAGAGAAAGATGAGGAGTTGTGAAAGAGAGGAGATG 1831  
 QY 1741 CTGATTAAGATCAAGAGAAAGATGAGAGCATGAGAAAGAGCATCAAGAGAGATA 1800  
 DB 1832 CTGATTAAGATCAAGAGAAAGATGAGAGCATGAGAAAGAGCATCAAGAGAGATA 1891  
 QY 1801 TTTGATCTGAGAGAAAGATTTTGTGAGGCTTTGAAAGAGCTCATGTAACAAGCATG 1860  
 DB 1892 TTTGATCTGAGAGAAAGATTTTGTGAGGCTTTGAAAGAGCTCATGTAACAAGCATG 1951  
 QY 1861 CACAAAGAGATGATGA 1878  
 DB 1952 CACAAAGAGATGATGA 1969

RESULT 4  
 AF239719 2254 bp DNA linear PLN 02-JUN-2000  
 LOCUS Arabidopsis thaliana SGS3 gene, complete cds.  
 DEFINITION AF239719  
 ACCESSION AF239719.1 GI:8164029  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE  
 AUTHORS 1 (bases 1 to 2254)  
 Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C.,  
 Morel, J.B., Jonette, D., Lacombe, A.M., Nikic, S., Picault, N.,  
 Remoue, K., Santal, M., Vo, T.A. and Vaucheret, H.  
 Arabidopsis SGS2 and SGS3 genes are required for  
 posttranscriptional gene silencing and natural virus resistance

CELL 101 (5), 533-542 (2000)

REFERENCE  
 AUTHORS Beclin, C., Mourrain, P., Vaucheret, H. and Elmayan, T.  
 Direct Submission  
 Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de  
 Saint-Cyr, Versailles 78026, France

FEATURES  
 source location/Qualifiers

1..2254  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="V"  
 /join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)  
 /product="SGS3"  
 /join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)  
 /codon\_start=1  
 /product="SGS3"  
 /protein\_id="AA073960.1"  
 /db\_xref="GI:8164030"  
 /translation="MSRAAGPMSKSKVAVGGRPEVAVOGLAGTFLASSDDGGEV  
 EVISKNKPKGNTSGKTWVQNSNPRAAGCOQGGSNVSGNNVSGRGNRG  
 IOANISGRGRASRYDNNFVAPVPVSPPLSGMNNQAGSGACHTAPFPDVED  
 VNASEENDSDALDDDDLDLDYDSDVSKSHGSKQKMFKFRGSLDLSIEO  
 INBPORWHPACONGPGAIDWYNLPLAARTKGARVRLHRELAEVLEKLOMKG  
 ASVTPGELTYGCKLIGDEKDEYIIVMPNVIIMTRLDKDDNKLWAGNQLLEF  
 DKTEALRASHSTGPOSHRMSIYMPRESSATYLEARHLRELALAMGLDRAMQGRM  
 FSGGVNLYGFLATKODLDTFNQHSQKTRLKFELKSGEVVVELKQISHDNOALNY

## ORIGIN

Query Match 77.8%; Score 1462; DB 8; Length 2254;  
 Best Local Similarity 83.3%; Pred. No. 0;  
 Matches 1878; Conservative 0; Mismatches 0; Indels 376; Gaps 4;

1 ATGAGTCTAGGGCTGTGCTCAATGCTTAAAGAAAAGAGGTCAGGGTGTATAGGCT 60  
 1 ATGAGTCTAGGGCTGTGCTCAATGCTTAAAGAAAAGAGGTCAGGGTGTATAGGCT 60  
 61 GAGGTGAACGTTGGTCAAGGTTTGGCAGGACGAGCTGGCTTCTTCAAGATGAT 120  
 61 GAGGTGAACGTTGGTCAAGGTTTGGCAGGACGAGCTGGCTTCTTCAAGATGAT 120  
 121 GGAGGAGAGTGGAGGTCATTTCCAGAGAACAGAACAAACGAGAAACATTTCTGA 180  
 121 GGAGGAGAGTGGAGGTCATTTCCAGAGAACAGAACAAACGAGAAACATTTCTGA 180  
 181 AAAAGTGGGTTCTCAGAAATTCGAATCTCTAGAGCTTGGGTGTCAAGCAAGG 240  
 181 AAAAGTGGGTTCTCAGAAATTCGAATCTCTAGAGCTTGGGTGTGTACGCAAGG 240  
 241 AGAGTAGACAGATATCTGGAGAGAAACAAATGTTCCGGAGAGAGTTACGGATGT 300  
 241 AGAGTAGACAGATATCTGGAGAGAAACAAATGTTCCGGAGAGAGTTACGGATGT 300  
 301 CGGGGCTTCAAGCTTAACTATCTGGTGGGAGCAGAGCTTGAAGCAAGATGATAC 360  
 301 CGGGGCTTCAAGCTTAACTATCTGGTGGGAGCAGAGCTTGAAGCAAGATGATAC 360  
 361 AACTTTGGCACCCCACTGTATCTGGCTCTCTTGGAGAGAGATGATGAG 420  
 361 AACTTTGGCACCCCACTGTATCTGGCTCTCTTGGAGAGAGATGATGAG 420  
 421 GCAAGAGAGGTTTGTGTGACACACAGCTGTGACAGAGTTCTCTGACGTGAGATG 480  
 421 GCAAGAGAGGTTTGTGTGACACACAGCTGTGACAGAGTTCTCTGACGTGAGATG 480  
 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTGAATCTGATGACGAC 540  
 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTGAATCTGATGACGAC 540  
 541 CTGGAGTGAATGATTAATGATCGGATGTGATCAAAAGACCATGATCAAGAAAGCAG 600  
 541 CTGGAGTGAATGATTAATGATCGGATGTGATCAAAAGACCATGATCAAGAAAGCAG 600  
 601 AATAAGTGTCAAAAAGATCTTTGGCAGCTTGGATGCTTGTGATCGAGCATTAAT 660  
 601 AATAAGTGTCAAAAAGATCTTTGGCAGCTTGGATGCTTGTGATCGAGCATTAAT 660  
 661 GAACCAAGAGGAGTGGATGCTCAGCTTGTCAAGACGAGCTGGTGCATGATGG 720  
 661 GAACCAAGAGGAGTGGATGCTCAGCTTGTCAAGACGAGCTGGTGCATGATGG 720  
 721 TATTAAGTGAACCTCTACTAGCTATGAGAGACAAAGAGCTGAGGCTTAAGCTC 780  
 721 TATTAAGTGAACCTCTACTAGCTATGAGAGACAAAGAGCTGAGGCTTAAGCTC 780  
 781 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGTCA 840  
 781 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGTCA 840  
 841 CTTGTGTGATGATTAATGAGGCAAGTGAAGAGGTTTGGTGAAGATTAAGATTAAGAA 900  
 841 CTTGTGTGATGATTAATGAGGCAAGTGAAGAGGTTTGGTGAAGATTAAGATTAAGAA 900  
 901 ATTGTCTGCTCCAAATGCTATCATATGATTAAGATTAAGATTAAGATTAAGATTAAG 960  
 901 ATTGTCTGCTCCAAATGCTATCATATGATTAAGATTAAGATTAAGATTAAGATTAAG 960

Oy	961	A-----	961
Db	961	AAAGGGAATTCCTCTGCTTTACTTCTTTAAATTTTCTCTGCATCTCACTGATCTTA	1020
Oy	962	-----AGTGGCTCGCGATGGGCAACCAAGCTGCTGAAATCTTCACAA	1007
Db	1021	GAATGTACATTTGATGGCTCGGATGGGCAACCAAGCTGCTGAAATCTTCACAA	1088
Oy	1008	GTAATGAGCTCTTAAAGCAGCGCCATTCTTATGCTCACAAGGCCATCTGTGGATGATGT	1067
Db	1081	GTATGAGGCTCTTAAAGCAGCGCCATTCTTATGCTCACAAGGCCCATCTGTGGATGATGT	1140
Oy	1068	TCTGATGTTTGAAGCAGCATGGCCTATCTTGAAGGCCAAGCCCTCCACCGGAGTT	1127
Db	1141	TCTGATGTTTGAAGCAGCATGGCCTATCTTGAAGGCCAAGCCCTCCACCGGAGTT	1200
Oy	1128	AGCTGAGATGGGGTTTAAAGATGAAATTGCGGGGTCAGAAAGCGCAGTATGTTTCTGGAG	1187
Db	1201	AGCTGAGATGGGGTTTAAAGATGAAATTGCGGGGTCAGAAAGCGCAGTATGTTTCTGGAG	1260
Oy	1188	TGTTTCGCAACGTATAGGCTTCCTTTCACCAAGCAGATCTGACATATTCATCA	1247
Db	1261	TGTTTCGCAACGTATAGGCTTCCTTTCACCAAGCAGATCTGACATATTCATCA	1320
Oy	1248	CTCTCA-----	1253
Db	1321	CTCTCAAGGTTCTCTCCCAAGAAATTTGATATATGCTTTTAAATTTGTCATTTGAAAT	1380
Oy	1254	-----AGG	1256
Db	1381	TTTAAAGTTTGGTCCGTGTTAATGATCTGTTATGATATATCTATGATTCATTTAG	1440
Oy	1257	CAAAACAAGCTGCAAAATTCAGATTGAATTCATACCAAGATGGTTGTAAGAGCTGAG	1316
Db	1441	CAAAACAAGCTGCAAAATTCAGATTGAATTCATACCAAGATGGTTGTAAGAGCTGAG	1500
Oy	1317	GCAGATCTCTGAGGCAATTCAGACGCTGAACTATTTAAGAACAAAGCTCTCAAAACGAA	1376
Db	1501	GCAGATCTCTGAGGCAATTCAGACGCTGAACTATTTAAGAACAAAGCTCTCAAAACGAA	1560
Oy	1377	CAAGACGCCAAGGCTTGAAGAAATCTCTGAAATTAATGAGCAGAAAGCTGCTAAGAC	1436
Db	1561	CAAGACGCCAAGGCTTGAAGAAATCTCTGAAATTAATGAGCAGAAAGCTGCTAAGAC	1620
Oy	1437	TGCAGAGGATATCGATCTGAGACAGAGAACTAAGATGAGCATGAAACGAACAGGGA	1496
Db	1621	TGCAGAGGATATCGATCTGAGACAGAGAACTAAGATGAGCATGAAACGAACAGGGA	1680
Oy	1497	AG-----	1498
Db	1681	AGAGGATATGATTTTCTAGAAATACAAACCTTGACATTTTGTATTACTAATGATTA	1740
Oy	1499	-----AGATGATGCAACGA	1514
Db	1741	CATTTTGATTAATATGTCAAACAAAACCTGTGTGTGTTGAAGATGATGACACGA	1800
Oy	1515	CAGGTTTTTTCATGATTCATCAACAGATCCATGAAGAGAGAGCGCAAGAGAGAGAA	1574
Db	1801	CAGGTTTTTTCATGATTCATCAACAGATCCATGAAGAGAGAGCGCAAGAGAGAGAA	1860
Oy	1575	TTTCGAGATGTTGACAGCAGAGAAAGTGCAGAGGTTGTTGGCCAGCAGCAGAGAAAT	1634
Db	1861	TTTCGAGATGTTGACAGCAGAGAAAGTGCAGAGGTTGTTGGCCAGCAGCAGAGAAAT	1920
Oy	1635	TAAATCCCTTAGCAATGACGATTTGCCGAAG-----	1665
Db	1921	TAAATCCCTTAGCAATGACGATTTGCCGAAGAGGTATATGTAATACTAATACATATCCCT	1980
Oy	1666	-----	1665
Db	1981	CTGGCGTTTTTTGTTTTCAAACCTAAGAGTAACTGAATTAATTCGGGTTTTGATTCCTTTCG	2040

OY		1666	AGAGCTGAGGAAGTGTCAACCTTCATCGAATTCCAGAAGAAAGATGTGGAGACTTTGT	1724
Db		2041	CAGAGCTGAGGAAGTGTCAACCTTCATCGAATTCCAGAAGAAAGATGTGGAGACTTTGT	2100
OY		1725	GGAAGAGAGGAGATGCTGATTAATAAGATCCAAGAGAAGATGGANAATGAAGAAG	1784
Db		2101	GGAAGAGAGGAGATGCTGATTAATAAGATCCAAGAGAAGATGGANAATGAAGAAG	2160
OY		1785	GCATCACAAGAGATTAATTGATCTCGAGAAAATTGATGAGCTTTGGAACAGCTCAT	1844
Db		2161	GCATCACAAGAGATTAATTGATCTCGAGAAAATTGATGAGCTTTGGAACAGCTCAT	2220
OY		1845	GTACAGAGTAGCCCTTCAATGAAGATGATGA	1878
Db		2221	GTACAGAGTAGCCCTTCAATGAAGATGATGA	2254
RESULT 5				
LOCUS	AX078760			
DEFINITION	AX078760	Sequence 1 from Patent WO0109591.	DNA	linear
ACCESSION	AX078760			
VERSION	AX078760.1	GI:13158379		
KEYWORDS				
SOURCE				
ORGANISM		Arabidopsis thaliana (thale cress)		
AUTHORS		Arabidopsis thaliana		
TITLE		Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
JOURNAL		Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;		
		rosids; euroside II; Brassicales; Brassicaceae; Arabidopsis.		
		Beclin,C., Rimayan,T. and Vaucherec,H.		
		Novel sgsl plant gene and use thereof		
		Patent: WO 0105951-A-1 25-JAN-2001;		
		AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE		
		AGRONOMIQUE (FR)		
FEATURES				
source		location/Qualifiers		
		1..3275		
		/organism="Arabidopsis thaliana"		
		/mol_type="unassigned DNA"		
		/db_xref="taxon:3702"		
	primer_bind	693..715		
		/note="p356AD/"		
	primer_bind	complement(2926..2952)		
		/note="p356Y"		
ORIGIN				
	Query Match	77.8%; Score 1462; DB 6; Length 3275;		
	Best Local Similarity	83.3%; Pred. No. 0;		
	Matches 1878; Conservative	0; Mismatches 0; Indels 376; Gaps 4;		
OY		1	ATGAGTCTCAGGGGTGGTCCAAATGCTTCAAGAAAAAGAACGTTCAAGGTGTTATAGGCT	60
Db		696	ATGAGTCTCAGGGGTGGTCCAAATGCTTCAAGAAAAAGAACGTTCAAGGTGTTATAGGCT	755
OY		61	GAGGTGAACAGTTGTGTTCAAAGTTTGCCAGAGGACGAGACTGCTTCTTCAAGATGAT	120
Db		756	GAGGTGAACAGTTGTGTTCAAAGTTTGCCAGAGGACGAGACTGCTTCTTCAAGATGAT	815
OY		121	GGAGGAAGTGGAGGATCTTCCAAAGAAACAAGAACCAACGAGAAACCTTCTCGA	180
Db		816	GGAGGAAGTGGAGGATCTTCCAAAGAAACAAGAACCAACGAGAAACCTTCTCGA	875
OY		181	AAAACCTGGGTTTTCTCAGAAATTCGAAATCTCTTGAAGCTTGGGGTGGTCAGACCAAGG	240
Db		876	AAAACCTGGGTTTTCTCAGAAATTCGAAATCTCTTGAAGCTTGGGGTGGTCAGACCAAGG	935
OY		241	AGAGTGACAACTATCTGGGAGAGGAAACAATGATTCGGGAGAGGTAAACGCAATGCT	300
Db		936	AGAGTGACAACTATCTGGGAGAGGAAACAATGATTCGGGAGAGGTAAACGCAATGCT	995
OY		301	CGGGGCAATTCAGCTTAATATCTCTGGGTCTGGGGAACGAGCGTTGAGCAAGAAATATATAC	360
Db		996	CGGGGCAATTCAGCTTAATATCTCTGGGTCTGGGGAACGAGCGTTGAGCAAGAAATATATAC	1055

[illegible]

OY	1257	CAAAA	CAAGGCTGA	AANTCG	AGTTGAA	ATCTAC	CAAGAA	TGCTGTAA	AGAGCTGAG	1316
Db	2136	CAAAA	CAAGGCTGA	AATTCG	AGTTGAA	ATCTAT	CAAGAA	TGCTGTAA	AGAGCTGAG	2195
OY	1317	GCAGAT	CTCTGAG	GACAA	TACG	CACCTG	GAATCT	TATGAA	CAAGCTCTCA	1376
Db	2196	GCAGAT	CTCTGAG	GAACAA	TACG	CACCTA	CTTATTA	GAAGCA	AGCTCTCA	2255
OY	1377	CAAGCA	CGCCAA	GGTCTT	GAGAA	TTCTTG	GAATTT	ATGAG	CGAAGAA	1346
Db	2256	CAAGCA	CGCCAA	GGTCTT	GAGAA	TTCTTG	GAATTT	ATGAG	CGAAGAA	2215
OY	1437	TGCAGAG	ATTAAT	TCGAT	CGTAGA	CAGGAA	CTAGAT	TGCA	GATGAACAA	1496
Db	2316	TGCAGAG	ATTAAT	TCGAT	CGTAGA	CAGGAA	CTAGAT	TGCA	GATGAACAA	2375
OY	1497	AG	-----	-----	-----	-----	-----	-----	-----	1498
Db	2376	AGAGGT	ATGATTTT	TCTAG	AAATCA	AAACTT	GACATTTT	GTATTA	CTACTG	2435
OY	1499	-----	-----	-----	-----	-----	-----	-----	-----	1514
Db	2436	CATTTT	GATTAAT	TATG	CCAA	CAAAAA	CTGTGT	GTGTTGA	AGATG	2495
OY	1515	CAGGTTT	TCATG	ATTCAT	CAATCA	AAACAG	ATCCAT	TGAAGA	AGACGCA	1574
Db	2496	CAGGTTT	TCATG	ATTCAT	CAATCA	AAACAG	ATCCAT	TGAAGA	AGACGCA	2555
OY	1575	TTTCG	ATGTTG	CAGC	ACG	ACGAA	CGTGC	CAAG	GTGTTGC	1634
Db	2556	TTTCG	ATGTTG	CAGC	ACG	ACGAA	CGTGC	CAAG	GTGTTGC	2615
OY	1635	TAA	TCCCT	TAG	CAATGA	CGAATT	GC	CGAAG	-----	1665
Db	2616	TAA	TCCCT	TAG	CAATGA	CGAATT	GC	CGAAG	-----	2675
OY	1666	-----	-----	-----	-----	-----	-----	-----	-----	1665
Db	2676	CTGCG	TTTTTT	GTTC	AACTA	AGTAA	CTGA	ATTA	TCGG	2735
OY	1666	-	AGAG	CTGA	AGAG	GTCA	AGCTT	CAAG	AAAGAA	1724
Db	2736	CAGAG	CTGA	AGAG	GTCA	AGCTT	CAAG	AAAGAA	AGATG	2795
OY	1725	GGAAG	AGAG	GAATG	CTG	TATA	AGAT	TAAGA	GAAGAA	1784
Db	2796	GGAAG	AGAG	GAATG	CTG	TATA	AGAT	TAAGA	GAAGAA	2855
OY	1785	GCAT	CAACA	GGAG	ATATT	TGAT	CTG	GAAGAA	GAATTTG	1844
Db	2856	GCAT	CAACA	GGAG	ATATT	TGAT	CTG	GAAGAA	GAATTTG	2915
OY	1845	GTAC	AAG	AGTGC	CTTCA	CAATGA	AGATG	ATTA	1878	
Db	2916	GTAC	AAG	AGTGC	CTTCA	CAATGA	AGATG	ATTA	1878	

RESULT 6	AB025633	AB025633	81365 bp	DNA	linear	PLN 27-DEC-2000
LOCUS	AB025633	Arabidopsis thaliana genomic DNA,				
DEFINITION	AB025633 BAO00015					
ACCESSION	AB025633.2	GI:10178221				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						



Sequence features of the regions of 3,076,755 bp covered by sixty p1 and TAC clones  
 DNA Res. 7 (1), 31-63 (2000)  
 MEDLINE  
 20181125  
 PUBLISHED  
 10718197  
 REFERENCE  
 2 (bases 1 to 81365)  
 Nakamura, Y.  
 Direct Submision  
 Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yama, Kiserazu, Chiba 292-0812, Japan (E-mail: ynakam@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)  
 On Sep 15, 2000 this sequence version replaced gi:4589439.  
 Address for correspondence: kaos@kazusa.or.jp  
 For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agg\_graph.cgi?c=M01  
 Genes with similarity to proteins in the databases are described in 'product', or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.  
 The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Graal-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).  
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
 This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K19M3 and the 3' clone is M011.

## FEATURES

## source

1. 81365  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 /clone\_id="M0M1"  
 /clone\_1fb="M1cnu1 P1"  
 join(3184..3403,3474..3595,3681..3723,3777..3898,4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)  
 /note="gb|AAF01580.1  
 gene\_id:M0M1.24  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97251.1"  
 /db\_xref="GI:8809710"  
 /translation="MHVQPGTSVDSPPAHKVDTPPKVQFLPISKTVDVDAQNPVS  
 YGSAFDDSSSNPLSPVFEPPSSFSFGGDDPLPGIEDLIQISGEPPGHEAQAG  
 YSINGTSCNFEWVCHLEBDGYNVYIDBSRETSNAGAKPNVLTADVVDLYLAIEV  
 OPLDNRNKGKFAFGKRYDVLVDYTLTFLVDFLSDCSVFEFMFALFPICYFLISG  
 ACEBDDMSQNIETKHTGTHASYSKSLAVGPVIMWAATLSIKREGVSTIKCISDLTAE  
 KESASTYTPRGPALVLIIGSDSHSRADNSPDLISGRDIEVTLRIFTRVM  
 FPKALNSFYHRCCHKSRGNL"  
 join(6317..6428,6508..6629,6714..6756,6841..6932,7121..7169,7396..7466,7550..7663,7785..7882,8112..8181)  
 /note="gb|AAF01580.1  
 gene\_id:M0M1.23  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97250.1"  
 /db\_xref="GI:8809709"  
 /translation="MOPGTSVSPALNOADPPISTKVVDTKNIFPDGDDAPL  
 PALBLOISGEPPGHEAQAGSYNGTSCNFEWVCHLEDSNVYIIGAKRPNTIYV  
 ADDVGLCAIEVQPLDNRNKGELVAVRANRDKACHPEQSNIDKLTGTHASYSK  
 SLAIGFVIMWAATLSIKREGVSTIKCISDLTAEKESASTYTPRGPALVLIIGSD  
 GSEHCLRVNDEPDLRLCKEERKVLVLYVVFQOKL"  
 complement(join(8642..8705,8794..8863,8978..9387,

## CDS

## CDS

## CDS

## CDS

## CDS

9584..10347))  
 /note="gene\_id:M0M1.22  
 unknown protein"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97249.1"  
 /db\_xref="GI:8809708"  
 /translation="MSMMKGSQWTAFDLKORQGLSEVEGDPFPVSTSVNASF  
 GVRGLRRNHSEPSKSSVLLPSPRPALITENKQCGOERGGCCRRKDDTLSPVNS  
 HDLAFYKIKENNSMADNDLIDVLTSTDDPEMALAFKGVSSGKDEDEPTSKI  
 SSNRSERYRTREKTVTSVKMAASTEDACKDILNSDGSFLVNASDNKFPDDI  
 SELDSIIIRLOSIPLEPMEBODLYLSRKQALKYRASNSHRAQVNAFOYDHASQ  
 KQHSKAREBDLMAKLAENAKKIIITGNNDNDITWKLDELHGLHATAVQAOERLQK  
 IEGHTRVRSVSPNRGKNALRSASQEPFRLDEBGMCHRTSRELRNLSLOVYTG  
 IGKHSRGASLPLAVKTFEEDRVRYPDETRPQVITVRPFRHS"  
 complement(12417..13424)  
 /note="contains similarity to unknown protein  
 gb|AAF27018.1  
 gene\_id:M0M1.21"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97248.1"  
 /db\_xref="GI:8809707"  
 /translation="MATDSQPNQKLTLPKTRIALTVISTMTDNAQRPDGTINRFLR  
 LFDFAPNPKPVNIVSTSDVFDQSDLMFELTPHYSGDKIPVVPFPGGFAPLS  
 PNAVYDPNCRFARKPAVYISVYRLAPERRYPADDDGALKYIEHNGSILPA  
 NADISRCFPAGDSAGNIAHVAIRICRSPSSFTAVKLGISTOPFGGERTBAE  
 KOLVAPLVSPPRTDMCKMGLNRDHAENVGVGNVADISLDPETRVVYAGDPDL  
 KDWQSYTEWMLCKRKATLIEYPMNFAFYFPLPEAGQILNRIKQVDRVASLS  
 A"  
 join(14922..14981,15072..15167,15563..15666,15755..15891,16096..16178)  
 /note="gene\_id:M0M1.20"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="50S ribosomal protein L24"  
 /protein\_id="BAA97247.1"  
 /db\_xref="GI:8809706"  
 /translation="MGKRAEKLIRHMKILNGDNVIMIRGDKGEGTIRYRSONR  
 VIVEGKLIKHKIKGPDHGGITFVEAEPHASNVQVDPYTGRCRKGVLYLEGTK  
 VRVARGTSGSIIIRPEILIKIRATPRFTGPKQTPMEFWEQTYDAKTGMDPL"  
 join(16823..16952,17143..17330,17412..17453,17675..17851,17943..18110,18199..18390)  
 /note="gene\_id:M0M1.19"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="26S proteasome, non-ATPase regulatory subunit"  
 /protein\_id="BAA97246.1"  
 /db\_xref="GI:8809705"  
 /translation="MERLORIYGAGGGIIGHASPDSPTLDTSEOVYISSLALMKLHG  
 RAGVMEYMGMLGEPDYEVTVVDVAVAMQSGVGVENVDYHPTNMLDMKQIG  
 RPEVNGVYHSHFGECWLSGVYDITQGSFELANRAVAVVDPIQSKGKVYDAFR  
 SINPTIMLGGEPRTGNTLNHLNPSIQALIHGNRYVSAIDNRYRNELEERKLN  
 LHKKQDGLTLRRDTSKTNIEQVQMLSLAAKYNXAVYSEBKLAIWVGR  
 ODAKKHLEHNSIMNSNIIVQTLGMLDPTVF"  
 complement(join(18965..19043,19240..19298,19419..19536,19666..19739,19908..19961,20387..20530))  
 /note="contains similarity to unknown protein  
 emb|CAA2897.1  
 gene\_id:M0M1.18"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97245.1"  
 /db\_xref="GI:8809704"  
 /translation="MDKNQQAIEKKMLVGNVEDEQQADESSLSFMEDLNRCAL  
 TTQKQFPAICLSGLTCTLLSMLVFPNPKFGITFTLGNLMALGSTAFLGPPOROV  
 TMLDPAIRIVATVYALIASIIIALFCALVVRNKLTLALIIIEFTGLIYVSYIIPAR  
 TMSKIEWTCDDTER"  
 join(12672..23634,23708..23999,24111..24355,24458..24624,24715..24925)  
 /note="emb|CAB62356.1  
 gene\_id:M0M1.17  
 similar to unknown protein"



		/codon_start=1 /evidence=not experimental /protein_id="BA97244.1" /db_xref="gi:8809703"		
		/translation="MSRPAQPMSEKRVYGGVYRVEQLVQGLAGRLASSODDGGEM EVLSKKNKKPNTSGTKTWSONSPFPAAGQOQGGKSNVSGKGNVSGDNGRG IOMNISGRALSRKRYDNNFVAPPEVPSFPLSGMWAQAGSAHQHTVAPPEVDE VINDASEENDSDALSDSDDLASDSDVSQKSHGSKRQKPKFRCVGLDLSLEO INPROMCIPACONGPDAIDWYNLHPLHARTKAGARVLAHRELAIVLEKLOMRG ASTPCGELYGOMKGLGEDEKDEIYVMPVLIINATRLDKDNDKMDRANQGLKSM DKREALRAHSHYSGPOSHRGMSYLAPESSATGVAERLHRELDKMDRANQGLKSM FSGVQOLYGLPLATKODLITPMOQKGTPLKELEKSYQEMVVELEQISEDDQOLNY FKRLSKONHAYLVLESLEIEMSEKRLTAEDNRLVRORTVQKQREEMDAHDEFF MDSIKLHERRDAKKEEFMLQOORAKYVQOQONINPSSNDCKRAEVGSFLIE OEHMEEFVEEREMLIKDEKEMDKRHHBEITDLKEKEDALDELQMLKHGLHND		
CDS	join(25614..25838,26061..26207,26324..26438,26516..26610, 26710..26813,27522..27765,27943..28036,28115..28305, 28385..28604,28688..28855,28944..29092) /note="gb AAC97420.1 gene_id="MQM1.16 strong similarity to unknown protein" /codon_start=1 /evidence=not experimental /protein_id="BA97243.1" /db_xref="gi:8809702"			
		/translation="MAPRAGQTAVAEAGADAPPOQGRFGSTISGIVIAVWFYF LRIASLALQKEMDPTASQIMTLFHKSGSLDWVFLSDEKFNFGDRLMYHETN ASPKFQKQKMDPTASQIMTLFHKSGSLDWVFLSDEKFNFGDRLMYHETN IPIVAVPEPESIRKSLTYTPSEPTLONNGSIYAHIPRAGSPDIPDEYQPLNSFR THAVATYPPKOKRKKKSLSGSPDSBPSEPRVKKDKDSDKKEVPEVEMISLWKP VTINLVDPFDLIVPPTGNAYPIITFNEMLRLDKPTIPYNETVSEPLNLEISP ISMKMOFLQVDQSPQWORYSGMLDESDLEKFLVLSNPFLIGITTFYFMSVSV DELAKNDIOFMNKKSMIEGLSAKSVLNTICQFVFLVLDLISWMLSSAGTGV IFPWKIGAMRIEVDKSGMIPRLRPHRESYVANKKEVDLDAIKRLSVLLVLLVGL STSYLAYERHKSRYKSMILSLTSCVWFGTLMCPOLFIYKVKVSAVLPMPQMTYKF LNTATIDLPATIKRPIIRLHLSVRDRIIFLIVQRMVYVVDKTVNEFVGSGDEET AEKKLITDEEEDKKTN" complement(join(29439..29576,29661..29885,29985..30268, 30341..30493,30579..30722,30956..31484)) /note="gene_id:MQM1.15		
Query Match	77.8%; Score 1462; DB 8; Length 81365;			
Best Local Similarity	83.3%; Pred. No. 0;			
Matches 1878; Conservative	0; Mismatches 0; Indels 376; Gaps 4;			
CDS				
1	ATGAGTTCTAGGCTGCTCAATGCTTAAGGAAAGAAAGCTTCAGGCTGTTATAGGCTT	60		
22672	ATGAGTTCTAGGCTGCTCAATGCTTAAGGAAAGAAAGCTTCAGGCTGTTATAGGCTT	22731		
QY	61	GAGTTGAA	CAGTTGTTGCAAGCTTTGGCAAGGACGAGCTGGCTTCTTCAACAAGATGAT	120
DB	22732	GAGTTGAA	CAGTTGTTGCAAGCTTTGGCAAGGACGAGCTGGCTTCTTCAACAAGATGAT	22791
QY	121	GGAAGAGAGTGGGAGGCTATTTCCAAAGAAACAAGAACAAACAAGAAACA	CTTTGCA	180
DB	22792	GGAAGAGAGTGGGAGGCTATTTCCAAAGAAACAAGAACAAACAAGAAACA	CTTTGCA	22851
QY	181	AAAACTTGGGTTTCTCAGAAATTCGAATCCTCTTCAAGCTTTGGGCTGCTCAGCAAGAGG	240	
DB	22852	AAAACTTGGGTTTCTCAGAAATTCGAATCCTCTTCAAGCTTTGGGCTGCTCAGCAAGAGG	22911	
QY	241	AGAGGTAAGCAAGTATCTGGGAGAGGAAACAATGATTCGGGAGAGGTAACGGCAATGAT	300	
DB	22912	AGAGGTAAGCAAGTATCTGGGAGAGGAAACAATGATTCGGGAGAGGTAACGGCAATGAT	22971	
QY	301	CGGGGCAATCAAGCTAATATCTGCTGGGAGAGAGCTTGAAGCAAGAAATGATTAAC	360	
DB	22972	CGGGGCAATCAAGCTAATATCTGCTGGGAGAGAGCTTGAAGCAAGAAATGATTAAC	23031	
QY	361	AACCTTGTGCAACCCCACTGATCTCGCCCTCTTGAAGAGAGATGGAATTTGCAAG	420	
DB	23032	AACCTTGTGCAACCCCACTGATCTCGCCCTCTTGAAGAGAGATGGAATTTGCAAG	23091	
QY	421	GGAAAGAGAGGTTCTGCTCAGACACACAGCTGTGAGAGAGTTTCTGATGACGAC	480	
DB	23092	GGAAAGAGAGGTTCTGCTCAGACACACAGCTGTGAGAGAGTTTCTGATGACGAGATGAT	23151	
QY	481	GGGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGATGATTTCTGATGACGAC	540	
DB	23152	GGGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGATGATTTCTGATGACGAC	23211	
QY	541	CTTGAAGATGATGATTAATGATCTCGAGTGTGATCAAAAGAGCCATGATCAAGAAACAG	600	
DB	23212	CTTGAAGATGATGATTAATGATCTCGAGTGTGATCAAAAGAGCCATGATCAAGAAACAG	23271	
QY	601	AATTAAGTGTCAAAAAGCTTTTGGAGCTTGAATGATCTTGTGATCGAGCAATTAAT	660	
DB	23272	AATTAAGTGTCAAAAAGCTTTTGGAGCTTGAATGATCTTGTGATCGAGCAATTAAT	23331	
QY	661	GAAACAAGAGAGTGTGATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT	720	
DB	23332	GAAACAAGAGAGTGTGATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT	23391	
QY	721	TATTAAGTGTCAAAAAGCTTTTGGAGCTTGAATGATCTTGTGATCGAGCAATTAAT	780	
DB	23392	TATTAAGTGTCAAAAAGCTTTTGGAGCTTGAATGATCTTGTGATCGAGCAATTAAT	23451	
QY	781	CATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	840	
DB	23452	CATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	23511	
QY	841	CTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	900	
DB	23512	CTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	23571	
QY	901	ATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	960	
DB	23572	ATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	23631	
QY	961	A-----	961	
DB	23632	AAGTGAATTCCTCTGCTTTTAACTTTTCTCTTGAATTTCTCTTGAATTTCTCTTGAATTTCT	23691	
QY	962	-----AGTGGCTGGGATGAGGACCAAGAGCTGTGAAATCTTGAACAA	1007	
DB	23692	GAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT	23751	
QY	1008	GATAGAGCTCTTGAAG	1067	
DB	23752	GATAGAGCTCTTGAAG	23811	
QY	1068	TCTGATGTTTGAAG	1127	
DB	23812	TCTGATGTTTGAAG	23871	
QY	1128	AGCTGATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1187	
DB	23872	AGCTGATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	23931	
QY	1188	TGTTGGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1247	
DB	23932	TGTTGGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	23991	
QY	1248	CTCTCA-----	1253	
DB	23992	CTCTCAAGGTTCTCTCCCAAGAAATTTGATATGATGATGATGATGATGATGATGATGATGATGAT	24051	
QY	1254	-----AGG	1256	
DB	24052	TTAAAGTTTGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	24111	
QY	1257	CAAAAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1316	
DB	24112	CAAAAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	24171	
QY	1317	GCAAGATCTGAGAGCAATAGAGAGCTGAATCTTAAAGAAAGAGCTCTCAAAAGAGAA	1376	

D	b		24112	GCAAGTCTGTGAGGACAAATCGACGCTGAACACTATTAAAGAACAAGCTCTCAAACAGAA	24233
Oy			1377	CAAGCAGCCCAAGTGCTTGAGGAATCTCTGGAAATTATGAGCGAAGAGCTGCTGAAC	1436
D	b		24232	CAAACAGCCCAAGTGCTTGAGGAATCTCTGGAAATTATGAGCGAAGAGCTGCTGAAC	24291
Oy			1437	TTCGAGAGATTAATCGGATCCGAGAACAGACAACTTAATCCACATGAAACAGAACGGA	1496
D	b		24292	TGCAGAGGATTAATCGATCTGTGAGACAGAGAACTTAAGATCACCATGAAACAGAGGA	24351
Oy			1497	AG-----AG-----	1498
D	b		24352	AGAGGTATGATTTTCTCTGAATAATCAAAAATTGACATTTGTATTACTACTGATTCA	24411
Oy			1499	-----AGATGATGACACAGA	1514
D	b		24412	CATTTTGTATTATATGTGCCAAAAAAACCTGTGTTGTTGTAAGTGTATCAACGA	24471
Oy			1515	CAGGTTTTTCATGGAATTCATCAACAGATTCATGAAGAAGACGCAAAAGAGAGAA	1574
D	b		24472	CAGGTTTTTCATGATTCATCAACAGATTCATGAAGAAGACGCAAAAGAGAGAA	24531
Oy			1575	TTTGAGAGTGTGAGCAGACAGAAAGCTGCAAGGTGTTGGCAGAGCAGAGAACAT	1634
D	b		24532	TTTGAGATGTTGAGCAGCAGAGAACGTGCAAGGTGTTGGCCAGCAGCAGAAACAT	24591
Oy			1635	TAAATCCCTTAGCAATGACGATTTGCCGAAG-----	1665
D	b		24592	TAAATCCCTTAGCAATGACGATTTGCCGAAGAGGTATATGTACTAATCAATATCCCT	24651
Oy			1666	-----	1665
D	b		24652	CTGCGCTTTTGTGTTTCAAACTAAGATGACTGATTAATTCGGTGTGATTTCTTGG	24711
Oy			1666	-AGAGCTGAGGAAGTGCACAGCTTCATCGAGTTTCAAGAGAGAGATGAGAGTTCT	1724
D	b		24712	CAGAGCTGAGGAAGTGCACAGCTTCATCGAGTTTCAAGAGAGAGATGAGAGTTCT	24771
Oy			1725	GGAAGAGAGGAGATGCTGATTAAGATCAAGAGAGAGAGATGAGAGACATGAAGAAG	1784
D	b		24772	GGAAGAGAGGAGATGCTGATTAAGATCAAGAGAGAGAGATGAGAGACATGAAGAAG	24831
Oy			1785	GCATCAGAGAGATTTGATCTGAGAGAAAGATTTGATGAGGCTTTGGAACAGCTCAT	1844
D	b		24832	GCATCAGAGAGATTTGATCTGAGAGAAAGATTTGATGAGGCTTTGGAACAGCTCAT	24891
Oy			1845	GTACACAGATGAGCTTCACATGAAGATGATTGA	1878
D	b		24892	GTACACAGATGAGCTTCACATGAAGATGATTGA	24925

RESULT 7  
ATHS28171

LOCUS  
Arabidopsis thaliana T-DNA flanking sequence, left border, clone 157C08.

ACCESSION  
AJ528171

VERSION  
AJ528171.1 GI:26796431

KEYWORDS  
left border; T-DNA flanking sequence.

SOURCE  
Arabidopsis thaliana (thale cress)

ORGANISM  
Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
1 Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepoint, L., Caboche, M. and Leclercq, A.  
T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites  
EMBO Rep. 3 (12), 1152-1157 (2002)

JOURNAL  
MEDLINE  
22363535

PUBMED	12446565
REFERENCE	2 (bases 1 to 650)
AUTHORS	Balzerque,S.
TITLE	Direct Submmission
JOURNAL	Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT	PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <a href="http://dbsgap.versailles.inra.fr/publications/">http://dbsgap.versailles.inra.fr/publications/</a> . This sequence has been generated in the framework of the French plant genomics program 'Genoplante' ( <a href="http://www.genoplante.com">http://www.genoplante.com</a> and <a href="http://genoplante-info.infobioingen.fr">http://genoplante-info.infobioingen.fr</a> ).
FEATURES	
SOURCE	Location/Qualifiers 1..650 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /cultivar="Wassiljewskij4" /db_xref="taxon:3702" /clone="157C08" /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
ORIGIN	misc.feature 1..650 /note="T-DNA flanking sequence left border"
Query Match	23.3%; Score 437.8; DB 8; Length 650;
Best Local Similarity	87.4%; Pred. No. 9e-96;
Matches 522; Conservative	0; Mismatches 2; Indels 73; Gaps 1;
Dn	733 CCTCTACAGGCTCATGGAGGACAAGAGGAGCTAGGCGAGTTAAGCTCCATAGAAGATTG 792
Dn	1 CCTCTACTAGCTCATGGAGGACAAGAGGAGCTTAGGCGAGTTAAGCTCCATAGAAGATTG 60
Dn	793 GCTGAAGTTTTAGAAAAGATCTTAAGATGAGAGCGCATCTGCATTCCTTGTTGTGTAG 852
Dn	61 GCTGAAGTTTTAGAAAAGATCTTAAGATGAGAGCGCATCTGCATTCCTTGTTGTGTAG 120
Dn	853 ATTTTAGGGCAGTGGAAAGGTTTGGTGGAGATGAAAGATTAAGAATTGCTCGGCT 912
Dn	121 ATTTTAGGGCAGTGGAAAGGTTTGGTGGAGATGAAAGATTAAGAATTGCTCGGCT 180
Dn	913 CCATGGTCATCATCAATGATATCTAAGCTGGATAAGACCATTAAGTA----- 961
Dn	181 CCAATGGTCATCATCAATGATATCTAAGCTGGATAAGACCATTAAGTAAGTGAATTC 240
Dn	962 ----- 961
Dn	241 TTCTGCTCTTAACTCTTTAAATTTTCTCTTGCAATCTCACTGATCTTAAGATGTTAACTT 300
Dn	962 --AGTGGCTCGGCAATGGGCAACCAAGAGCTGTGAAATACCTGCACAAGATGAGGCTCT 1019
Dn	301 GTAGTGGCTCGGCAATGGGCAACCAAGAGCTGTGAAATACCTGCACAAGATGAGGCTCT 360
Dn	1020 TAGAGCACGCCATTCCTATATGTCTCAKAGGGCCATCGTGGGATATGATGTTCTTGATGTTTGA 10797
Dn	361 TAGAGCACGCCATTCCTATATGTCTCAKAGGGCCATCGTGGGATATGATGTTCTTGATGTTTGA 420
Dn	1080 GAGCAGTGCACATGGCTATTTTGGAGGCGGAAGCGCTCCACCGGGAAGTTAGCTGAGATGGG 11399
Dn	421 GAGCAGTGCACATGGCTATTTTGGAGGCGGAAGCGCTCCACCGGGAAGTTAGCTGAGATGGG 480
Dn	1140 GTTAGATAGAAATTCCTGGGGGTGAAGCGCAGATATTTTCTTGAGAGTGTTTCGCCAACT 11999
Dn	481 GTTAGATAGAAATTCCTGGGGGTGAAGCGCAGATATTTTCTTGAGAGTGTTTCGCCAACT 540
Dn	1200 GTATGAGCTTCTCTTGCAACGAAGCAAGTCTGAGATATTTCAATTAACAATCTCTCAAG 1256
Dn	541 GTATGAGCTTCTCTTGCAACGAAGCAAGTCTGAGATATTTCAATTAACAATCTCTCAAG 597



OY 1057 GGGATGAGTGTCTGATGTTTGAAGACAGTGGCCATGCTATTGGAGGCCGAGACGCTC 1116  
 |||||  
 Db 1070 GGCAATGAGTGTCTATCTTTGACGCTGGCTGTGCTATATGAGACAGAACCTTTG 1129  
 |||||  
 OY 1117 CACCGGAGTGTAGTGTGAGTGGGTTAGATAGATATGCTGGGT-----CAGAGCG 1169  
 |||||  
 Db 1130 CATGATCATCTTTGTTGCTCAAGAACAGACAGGAATATCATGAAACAGTCTCCACCAAGT 1189  
 |||||  
 OY 1170 CAGTATGTTTCTGAGAGTGTTCGCCAATGATGCTTCCTTGCACGACGACGATCT 1229  
 |||||  
 Db 1190 TACATCTTACCTGCTGGGAGAGAGAGCTGTATGCTTTCTTACCTAACAAAGATACAT 1249  
 |||||  
 OY 1230 GGAACATATTCATCAACACTCTCAAGAGCAAAACAAAGCTGAAATTCAGATTGAAATCAT 1289  
 |||||  
 Db 1250 GGAGCATTCATTAATGACATTCGCCACGAGAAATCCCTTTGAATATGAGATGATCTTA 1309  
 |||||  
 OY 1290 CCAAGATGATGTTGTAAAGAGCTGAGGAGATCTCTGAGGACAAATCAGCAGCTGAAT 1349  
 |||||  
 Db 1310 TAATGATGATGTTGTTTACCAATATGAGAGATGAGAGCAACCAACCACTGAAATTA 1369  
 |||||  
 OY 1350 CTTTAAAGAACAGCTCTCTCAAAACAGAACAGACCGCAAGCTCTTGAAGATCTCTGA 1409  
 |||||  
 Db 1370 CCTGAAGAACAAATGTTTAAAGAGAGAGAGCTCTTAATTAGTAGAGATATACATTGAG 1429  
 |||||  
 OY 1410 AATTATGAGCGAAGAGCTGCTAGAACCTGACAGATATTCGATCTGAGACAGAGAAC 1469  
 |||||  
 Db 1430 TCTTTTACCCAGAAACCTTAAAGAGAGACATGAGAGAAATCTATTGTACGAAATTAAC 1489  
 |||||  
 OY 1470 TAAAGTGCAGCATGAAACAGAACAGAGAGATGATGACACAGCAGGTTTTTCATGGA 1529  
 |||||  
 Db 1490 CAAAGAAAAAGCACTTGAGATATGAGAAAGATGAAATGACAGAGAGAAATTTTTCATGA 1549  
 |||||  
 OY 1530 TTCAATTC-AAACAGATCCATGAAAGAGAGAGAGAGAGAAATTTTCAGATGTTTC 1588  
 |||||  
 Db 1550 TCAATTCGAAAAAATTCACAAAGCCACAGAGAGAGAAATTCAGATTGAGAAATGTC 1609  
 |||||  
 OY 1589 AGCAGCAGAGAGCTGCGCAAGGTTGTTGCGCAGAGAGAGAGAGAAATTTTCCTCTAGCA 1648  
 |||||  
 Db 1610 TGCAG 1660  
 |||||  
 OY 1649 ATGACGATTTCCGAAAG 1708  
 |||||  
 Db 1661 CTGAAGATGCGAGGAG 1720  
 |||||  
 OY 1709 AGATGAGAGAGATTTTGGAGAGAGAGAGAGAGAGATGCTATTAAGATCAAGAGAGAGATG 1768  
 |||||  
 Db 1721 ATGTTGAGAGATTTTAAAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1780  
 |||||  
 OY 1769 AAGACATGAAAGAGAGAGAGAGAGAGAGAGAGAGATTTTGTGAGAGAGAGAGAGAG 1828  
 |||||  
 Db 1781 TGAAGCTCAAG 1840  
 |||||  
 OY 1829 CTTTGAAGAGCTCATGTATCAAGCA 1853  
 |||||  
 Db 1841 CCTCAGCTTCCCTGATGATTAACA 1865  
 |||||

RESULT 9  
 AKI00699 2570 bp mRNA linear PLN 24-JUL-2003  
 LOCUS AKI00699  
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J023114M02, full  
 insert sequence.  
 ACCESSION AKI00699  
 VERSION AKI00699.1 GI:32985908  
 KEYWORDS FLI\_CDNA; CAP tripper.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Euphorbiales; Oryzaceae; Oryza.  
 REFERENCE 1  
 AUTHORS The Rice Full-Length cDNA Consortium, National Institute of

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 COMMENT

Agrobiological Sciences Rice Full-length cDNA Project Team:  
 Kikuchi, S., Sato, K., Nagata, T., Kawagashira, N., Doi, K.,  
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,  
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,  
 Ohtsuki, K., Shishiki, T., Foundation of Advancement of International  
 Science Genome Sequencing & Analysis Group: Otsu, Y., Murakami, K.,  
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,  
 Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,  
 Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,  
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,  
 Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,  
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arikawa, T., Fukuda, S.,  
 Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,  
 Kagawa, I., Kondo, S., Komori, H., Miyazaki, A., Otsu, N., Ota, Y.,  
 Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T.,  
 Yoshino, M., and Hayashizaki, Y.  
 Collection, mapping, and annotation of over 28,000 cDNA clones from  
 japonica rice  
 Science 301 (5631), 376-379 (2003)  
 22752273  
 12869764  
 2 (bases 1 to 2570)  
 Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Doi, K.,  
 Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,  
 Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,  
 Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,  
 Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,  
 Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,  
 Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Komori, H., Kouda, M.,  
 Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,  
 Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,  
 Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,  
 Namiki, T., Narikawa, R., Niikura, J., Niishi, K., Nomura, K.,  
 Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,  
 Otsu, N., Ota, Y., Otsu, Y., Otsu, Y., Ryo, R., Saitoh, H., Sakai, C., Sakai, K.,  
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Sato, K., Shibata, K.,  
 Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,  
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,  
 Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,  
 Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,  
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and  
 Yoshimura, A.  
 Direct Submission  
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of  
 Agrobiological Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail:skikuchi@ribs.affrc.go.jp,  
 Tel:81-29-838-7007, Fax:81-29-838-7007)  
 This clone is one of the 28k full-length cDNA clones from japonica  
 rice.  
 URL: <http://cdna01.dna.affrc.go.jp/cDNA/>  
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Sato, K.,  
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., and  
 Yamamoto, M.  
 PAYS genome Sequencing & Analysis Group: Otsu, Y., Iida, Y.,  
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,  
 Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,  
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
 Yoshimura, A., Matsubara, K., and Murakami, K.  
 Genome Exploration Research Group in Riken Genomic Sciences Center  
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
 Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,  
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Komori, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Niishi, K., Nomura, K., Numasaki, R., Ohno, M., Otsu, N.,  
 Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,

Saeki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagami, T., Tanaka, Y., Tagawa, A., Takahashi, F.,  
Takaku-Akashita, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,  
Yasunishi, A. and Hayashizaki, Y.  
Location/Qualifiers

FEATURES  
source 1..2570  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultiar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="U023114M02"

## ORIGIN

Query Match 19.8%; Score 372; DB 8; Length 2570;  
Best Local Similarity 57.0%; Pred. No. 1e-79;  
Matches 786; Conservative 0; Mismatches 575; Indels 19; Gaps 5;

477 TGATGTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTGTGATGATTTCTGATGA 536  
713 TGAATGTGACAAATGTGAGAGATTAATGATTCATGATGATGATGATGATGATGATGATGAT 772  
537 CGACCTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 596  
773 TGATGATACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 832  
597 GCAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 656  
833 AAGCACAAGCTGTGAGAGAGATTTTGAATCTTGAAGAGATTAAGTGTGAACAGAT 892  
657 AATGAGACCAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 716  
893 AATGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951  
717 TTGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 773  
952 CTGTATCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011  
774 TAAAGTCAATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 833  
1012 TAAAGTCAATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071  
834 TGTATTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 893  
1072 AATGATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1128  
894 TTAATGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 953  
1129 TGTATTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1188  
954 TAAAGTCAATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1013  
1189 GATATGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1248  
1014 GGTCTTGAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1073  
1249 CGTGAAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1308  
1074 GTTGAAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1133  
1309 ATTGAAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1368  
1134 GATGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1193  
1369 TCAAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1428  
1194 CCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1253  
1429 GATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1485  
1254 AGGCAAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1313  
1486 TGAAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1545

1314 GAGCAGATCTCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1373  
1546 GAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1605  
1374 GAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1433  
1606 AGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665  
1434 AACTGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1493  
1666 AGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1725  
1494 GGAAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1553  
1726 GGAAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1785  
1554 AAGAGAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1613  
1786 CACGAGAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1845  
1614 TGCCAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1673  
1846 AG-----AGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1896  
1674 GGAAGTCAAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1733  
1897 AGAGATCAAGAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1956  
1734 GGAAGTCAAGAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1793  
1957 AGAGATCAAGAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2016  
1794 GGAAGTCAAGAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1853  
2017 AAGAGAT 2076

## RESULT 10

AK064217 1556 bp mRNA linear PLN 24-JUL-2003  
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:002-104-F07, full  
DEFINITION insert sequence.

ACCESSION AK064217.1 GI:32974235  
VERSION AK064217.1  
KEYWORDS FLI CDNA; oligo-capping.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

## REFERENCE

1 The Rice Full-Length cDNA Consortium, National Institute of  
Agricultural Sciences Rice Full-Length cDNA Project Team,  
Rikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,  
Kishimoto, N., Yasaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,  
Kojima, K., Namiki, T., Ohneda, E., Yanagi, M., Suzuki, K., Li, C.,  
Ohtsuki, K., Shishiki, T., Foundation of Advancement of International  
Science Genome Sequencing & Analysis Group; Otsu, Y., Tsunoda, Y.,  
Iida, Y., Sugano, S., Fujimura, T., Maeda, H., Kobayashi, M., Xie, Q., Lu, M.,  
Kurosaki, T., Kodama, T., Mizuno, K., Yokomizo, S., Nishikura, J.,  
Nakawara, R., Sugiyama, A., Kawamata, M., Yoshimura, A., Miura, J.,  
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., MIURA, J.,  
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,  
Kawai, J., Carninci, P., Adachi, U., Aizawa, K., Arakawa, T., Fukuda, S.,  
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,  
Kagawa, I., Kondo, S., Kono, H., Miyazaki, A., Otsu, N., Ota, Y.,  
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Yoshino, M., and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from  
japonica rice  
Science 301 (5631), 376-379 (2003)  
JOURNAL MEDLINE  
22752273  
PUBMED 12869764



REFERENCE  
2 (bases 1 to 1556)  
Aadachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Heshizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirotsu, T., Hori, F., Hotte, I., Iida, J., Iida, Y., Ikeda, R., Imanura, K., Imocani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koye, S., Kurihara, C., Kurotsaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osaio, N., Ota, Y., Otsomo, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shigenaga, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

TITLE  
Journal  
Direct Submission  
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT  
This clone is one of the 28K full-length cDNA clones from japonica rice.  
URL : [http://cdna01.dna.affrc.go.jp/cDNA/NIAS\\_Rice\\_Full-Length\\_cDNA\\_Project\\_Team:Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotte,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.andYamamoto,M.](http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-Length_cDNA_Project_Team:Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotte,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.andYamamoto,M.)

FAIS Genome Sequencing & Analysis Group: Otsomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurotsaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nishikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken:  
Aadachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Heshizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirotsu, T., Hori, F., Iida, Y., Imanura, K., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohta, N., Ota, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saeki, D., Sato, K., Shibata, K., Shigenaga, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES  
source  
1..1556  
Location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultiar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="002-104-F07"

ORIGIN  
Query Match 19.6%; Score 368; DB 8; Length 1556;  
Best Local Similarity 60.7%; Pred. No. 9,6e-79;  
Matches 659; Conservative 0; Mismatches 415; Indels 12; Gaps 3;  
460 TTCTCTGAGTGGAGATGATGATATGCTTCTGAGGAGAGAGATTCGATGCT 519  
473 TTGCTGTTGTTGATCTGAAACAGACATGCTGAAGGTTATGACATCGACATGAT 532

QY 520 TTGATGATTTCTGATGACGACCTTGCAAGTGAATTATGACTCGATGTCGACAAAG 579  
DB 533 AATGATGATGATATGATGACATTTGATGATGACTATGATTCGATGACGAGGAAA 592  
QY 580 ACCCATGATGATCGAAGAGCAAAATAGTGTTCAAAAAGTCTTTGGCAGCTTGATAC 639  
DB 593 AATTTTGAACATCGGAAAAAATCAAGATTGTTCAAGGTTCTTTCGAACTCTGAGGGC 652  
QY 640 TTGTCGATGACGACGATTAATGAACACAGACGAGTGGCACTTGTCCAGCTTGAGAC 699  
DB 653 TTAAGTGTGAACAGTTAATTAATACCACTAGGCAATGGCACTTGGCCAGATGCAAAAT 712  
QY 700 GGACCTGTGGCCATGATTTGATTA--CTGCACTCTTACTAGCTCATGCGAGACA 756  
DB 713 GGACCTGTGGCAATTTGATGATTAAGGGCTGACAACTTTGATGATGATGCAAAACA 772  
QY 757 AAAGAGCTGAGCGAGTTAAGTCCATGAGAAATTTGCTGAAGTTTGAAGAAATCTA 816  
DB 773 AAGGTTCTTAAAGTCAAGCTGACAGAAATTTGCTAGTTTATTTGAAAGAGCTA 832  
QY 817 CAGATGAGAGCGCATCTGTCACTTCTGTGTGAGATTTATGGCAGTGAAGGTTTG 876  
DB 833 TTCTCGAGGGAACTTCAGTGTACATCTGCGGCAATTTAGAAATGGAAGGATTG 892  
QY 877 GGTGAGATGAAAGATTAATTAATTTGCTGCTCCATAGTGCATCATGAAATCT 936  
DB 893 CG---GGAAGCATGATGATGATGATTTGTTGGCCACAAATGTTGTTTGAACAT 949  
QY 937 AGACTGATTAAGACGATTAAGATGATGCTGCGCATGCGCAACGAAAGCTGCGAA 996  
DB 950 GTATTGAAACAGATGAGGATGATTAAGTGAAGGTAATGCGGAAACCAAGACTAATGAT 1009  
QY 997 TACTTCGACAGTATGAGGCTTTGAGACGCGCACTTCTATGTCGCAAGGCCATGCT 1056  
DB 1010 TATTTCACTGAATATGCTGCAAGCAAGCAGCCATGCTTATGTCGCAATGCGACCT 1069  
QY 1057 GGGATGAGTGTCTGATGTTTGAAGAGAGTGCCTGCTATTTGAGGCGCAAGCGCTC 1116  
DB 1070 GGCATGATGATGATCTTATCTTGAACGCTCGCTGCGCTATATGGAAGCAGAGCTTG 1129  
QY 1117 CACCGGAGTTAGCTGAGATGGGTTAGATGAAATTCCTGG-----GGTCGAAGCGC 1170  
DB 1130 CATGATCACTTTGTTGTCGCAAAAGACAGAGATTAATGAGACAGTCTCACAAGTT 1189  
QY 1171 AGTATGTTTTCGAGAGTGTTCGCCAATGATGCTCTTGCACAGCAAGCAATCTG 1230  
DB 1190 ACATTTCTTACCGGAGGAGAAAGGAGCTGATGTTCTTATGTCGCAAAAGATGACATG 1249  
QY 1231 GACATATTCATCAACCTCTCAAGGCAAAACAAAGCTGAATTTGAGTGAATCATATC 1290  
DB 1250 GAGACATTCATATGACATTTGCCAGGAAATTCGTTGAAATATGATGATGATCTTAT 1309  
QY 1291 CAAAGATGTTTGAAGAGAGCTGAGGAGATCTTGAAGACATCAGACGCTGAATCTAC 1350  
DB 1310 AATGATGATGATGATACCAATGAAAGATGAGCAAGCAACACAGAACTGAATATC 1369  
QY 1351 TTTAAGAACAGCTCTCAAAACAGAACAGCAGCAGGCTGTTGAGGAATCTCGAA 1410  
DB 1370 CTGAAGAACAAATGTTTGAAGAAAGAGCAGCTTAAATTTAGTGAAGATTCATTTGAT 1429  
QY 1411 ATTATGAGAGAGAGCTGATGAACCTCAGAGGATTAATCGATGTTGAGCAGAGAACT 1470  
DB 1430 GTTGTACCCGAAACTTATGAGAGACAAATGAAAGAAATCATTTGTACGAAATTAAGCC 1489  
QY 1471 AAGATGACAGTGAACAGAACAGGAGAGATGATGACACAGCAGGTTTTTCATGAT 1530  
DB 1490 AAAAGAAAGCACTTGAATGAGAAAGATGAAGATGACGAGAGAAATTTTTCATGAT 1549  
QY 1531 TCAATC 1536  
DB 1550 CAATC 1555

RESULT 11  
AF542974 1564 bp mRNA linear PLN 02-JUL-2003  
LOCUS Trilicium aestivum Emr1 mRNA, complete cds.  
DEFINITION AF542974  
ACCESSION AF542974.1 GI:32401385  
VERSION  
KEYWORDS  
SOURCE Trilicium aestivum (bread wheat)  
ORGANISM Trilicium aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1564)  
Zhao, X., Li, Q. and Zhang, X.  
Isolation and expression of a new kind of gene involve in  
embryogenesis in Trilicium aestivum L.  
REFERENCE  
AUTHORS Unpublished  
TITLE 2 (bases 1 to 1564)  
JOURNAL  
REFERENCE Zhao, X., Li, Q. and Zhang, X.  
AUTHORS Direct Submission  
TITLE Submitted (03-SEP-2002) College of Life Sciences, Shandong  
JOURNAL Agricultural University, Daizong Street, Tai'an, Shandong 271018,  
P.R. China  
Location/Qualifiers  
FEATURES  
Source 1..1564  
/organism="Trilicium aestivum"  
/mol\_type="mRNA"  
/cuiivar="PH 82-2-2"  
/db\_xref="taxon:4565"  
/dev\_stage="embryo"  
309..1223  
/note="involved in embryogenesis"  
/codon\_start=1  
/product="Emr1"  
/protein\_id="AAP80862.1"  
/db\_xref="gi:32401386"  
translation="MGNOELLETFSDYATKAAHAYPGHGMVILIESSAVGYME  
ARLHKPIIDORTDQNRNRPFLPGKRLQYGFARKEDETRRHQCGKSRKY  
EMRSHNEVAVOMKQSEMDNQLNLYLNKRVKTEORSKVEETLVGYTQKRTMEN  
IFVRKAKEXSEVEEEMKSGKEFHQIENIHAKTIDKSEFEERLQGERARQCD  
VDSGTENRRLRKEVOYRFTFCQYKNDQVEFAEDDEMIKAEKRYQLKKEYAKAVE  
LEKFPDALTGLMEKRPCTQASSSSP"  
ORIGIN  
Query Match 15.5%; Score 291.6; DB 8; Length 1564;  
Best Local Similarity 59.2%; Pred. No. 4.2e-60;  
Matches 539; Conservative 0; Mismatches 359; Indels 12; Gaps 2;  
OY 952 GATTAACGATAGTGGCTGGCATGGCAACCAAGAGCTGCTGATTAAGTATGACAAAGTAT 1011  
DB 288 GATGCTCAGTATGCTGTTAGGACGAGCAACCAAGACTTCTTGAATATTTCAAGTATAT 347  
OY 1012 GAGGCTTTAGAGACAGCCATTCTATGCTCCACAGGCGCATGCGGATGATGTTCTG 1071  
DB 348 GCTGCGACCAAGACAGCCATGATATGATGCGGACCGCTGCGCATGATGCTGA 407  
OY 1072 ATGTTTGAAGCACTGCTGCTATTTGAGAGCCGAGCGCTCCACCGAGATTAGCT 1131  
DB 408 ATATTGAAAGCTCTGCTGCGGTATATGAGAGCAAGAGCTTCTTAAGCATTTTAT 467  
OY 1132 GAGATGGGGTTAGATAGATTTGCTGCTG---GGGTCAAGAGCGAGTATGTTTCTGGAGT 1188  
DB 468 GATCAAAAGACAGACAGGACACTTGGCAAGATGCGAGGTTCTTTTCTTACCTGTTG 527  
OY 1189 GTTCCGCAACTGTATGCTTCTTGGCAACGAGCAATCTGACATATTCAATCAAC 1248  
DB 528 AAGAGACATTAATACGCTTCTTCCGCAAGAGAGCAATGAGCATTTCAATAGAC 587  
OY 1249 TCTCAAGCAAAACAGGCTGAAATTCAGATTGAATCATACCAAGATGTTGTAAG 1308  
DB 588 TGCCAGAGGAAAAGCCGCTGAAATACAGATGATCTCATATGAGATGATGCGCC 647  
OY 1309 GAGCTGAGGCAAGTCTTGAAGACATGACGCTGAACCTTAAGAACAGCTCTCA 1368

DB 648 CAGATGAGCAAGTAATGACCGAAGTAAATCAACAACTCAATTACTGTAAGAAACAGGCTGTT 707  
OY 1369 AAACAGAAACAAGACAGCCAGGCTGTTAGAAATCTCTGAAATTTGAGCGCAAGCTG 1428  
DB 708 AAGACAGACAGCGCTCTAAAGTTGTAGAAAGAAACCTGGGTGTATTTACGAGAACTT 767  
OY 1429 CGTAGAATCTGACAGATTAATCGATATGCTGAGACAGAACTAAGATGACATGACAG 1488  
DB 768 CGAGAGCTATGAGAGAAATATATTGCTGACAGATTAACCTAAGAGAAAGCACTCTGAG 827  
OY 1489 AACAGAGAGATGATGACACAGACAGTTTTCATGATTCATCAATCAAGATTCAT 1548  
DB 828 TATGAGAGAGATTAATCCACAGAAAATTTCTTCATATCATGATGAGAACATTTAC 887  
OY 1549 GAAAGAGAGACGCAAGACAGAGAAATTTGAGATGTTGACGACGACGAGACGTCACAG 1608  
DB 888 AAGGCCACAGAAAGACAGAGAAAGCGATTGAGATTCCTGACGAGAGAGCGTGCAG 947  
OY 1609 GTTGTGGCCAGACAGACAGAACATTAATCCTCTAGCAATGACATTTGCCAAAGAGA 1668  
DB 948 GCT-----AGACAGTGTATGATGATTTCTGGAGTCTGAAATGCGAGCTAAG 998  
OY 1669 GCTGAGAAAGTGTCAAGCTTCTCATCGATTTCAAGAAAGATGAGAGAGTTTGTGAA 1728  
DB 999 AAGAACAGGTACAGCGGTTCTATGATGCTCAGGTTAAGACCTGACAGAGTTGAGGCT 1058  
OY 1729 GAGAGAGATGCTGATTAAGATCAAGAAAGAAAGATGAAAGACATGAAAGAGGAT 1788  
DB 1059 GAAAGGACCAATGATTAAGCGGACGAGAGAAAGTGCACACTGAAGAAAGATGAC 1118  
OY 1789 CACGAGAGATATTGATCTGAGAAAGATTTGATGAGGCTTTGAAACAGCTCATGAC 1848  
DB 1119 ATGCGAAAGAGGAGGCTCGAAGAGATTGATGCTGCTCACCAGCTTTGATGAG 1178  
OY 1849 AAGCATGGCC 1858  
DB 1179 AAACATTGGC 1188

RESULT 12  
AK061975 1214 bp mRNA linear PLN 24-JUL-2003  
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:001-043-A02, full  
DEFINITION  
insert sequence.  
ACCESSION AK061975  
VERSION AK061975.1 GI:32971993  
KEYWORDS FLI CDNA; oligo-capping.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriatoidae; Oryzaceae; Oryza.  
1  
The Rice Full-length cDNA Consortium, National Institute of  
Agricultural Sciences Rice Full-length cDNA Project Team;  
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,  
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,  
Kojima, K., Nanki, T., Ohneda, E., Yahagi, M., Suzuki, K., Li, C.,  
Kobayashi, K., Shishiki, T., Foundation of Advancement of International  
Science Genome Sequencing & Analysis Group; Oono, Y., Murakami, K.,  
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tanoda, Y.,  
Kurosaki, R., Kodama, T., Maeda, H., Kobayashi, M., Xie, Q., Lu, M.,  
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Miura, J.,  
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., RIKEN,  
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,  
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arikawa, T., Fukuda, S.,  
Hara, A., Hashidume, M., Hayatsu, N., Imetani, K., Ishii, Y., Itoh, M.,  
Kagawa, I., Kondo, S., Komori, H., Miyazaki, A., Otsu, N., Ota, Y.,  
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T.,  
Yoshino, M., and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from  
japonica rice



JOURNAL Science 301 (5631), 376-379 (2003)  
 MEDLINE 22752273  
 PUBMED 12869764  
 REFERENCE 2 (bases 1 to 1214)  
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imocani, K., Iehibiki, J., Iehi, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyu, S., Kurihara, C., Kurotsaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, K., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, M., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28k full-length cDNA clones from japonica rice.  
 URL: [http://cdna01.dna.affrc.go.jp/cDNA/NIAS\\_Rice\\_Full-length\\_cDNA\\_Project\\_Team:kikuchi,S.,Sato,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yahagi,M.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.andYamamoto,M.](http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project_Team:kikuchi,S.,Sato,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yahagi,M.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.andYamamoto,M.)

PAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Iehibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurotsaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nishikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Imamura, K., Imocani, K., Iehi, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shishiki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES  
 source location/Qualifiers  
 1..1214  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /culivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="001-043-A02"

ORIGIN  
 Query Match 15.4%; Score 288.8; DB 8; Length 1214;  
 Best Local Similarity 58.7%; Pred. No. 2e-59;  
 Matches 544; Conservative 0; Mismatches 367; Indels 15; Gaps 2;  
 934 ACTAGACTGATAGAGACGATTAAGTGTCTCGGATGCGCAACCAAGACTGTCTG 993

Db 2 ACTGATTTGGAACAAGATGAGATGATTAAGTGAAGGATATGCGGGAACCAAGACTAAAT 61  
 994 GAATCTTCGACAAAGATATGAGGCTCTTGAGACGCGCATTCCTATGTGTCACAGGCGCAT 1053  
 Db 62 GATTTATTTCAAGTAATATGCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 121  
 1054 CCGTGGATGAGTGTCTGATGTTTGAAGACAGTGCACCTGCTATTTGGAGGCCGCAAGC 1113  
 Db 122 CGTGACATGAGTGTCTATCTTTGACAGCTCGGCTGTGAGGCTATATGAGACAGACCT 181  
 1114 CTCACCGGAGTATGCTGAGATGCGGTTAGATGAATTCCTCG-----GCTCAGAG 1167  
 Db 182 TTGCATGATCACTTTGTTGCTCAAGAAACAGACAGGAATACATGCAACAGTCTCACAAG 241  
 1168 CGCAGTATCTTTCTGAGAGTGTGGCCACCTGTATGCTTCTTGCAACGAAAGAT 1227  
 Db 242 GTTACATTTCTTACCTGTGGGAAAGGCAAGCTGTATGTTCTTAGCTACAAAGATGAC 301  
 1228 CTGACATATTCATCAACACTCTCAAGGCAAAACAAGGCTGAATTCAGTTGAATCA 1287  
 Db 302 ATGAGACATTCATATGACATTGCCACGAAATCCCGTTGAAATATGAGATGAGATCT 361  
 1288 TACCAAGATGCTGTGAAGAGGCTGAGCAGATCTCTGAGACATCAGACCTGAAC 1347  
 Db 362 TATATGATGATGTGTGTTTACCAATGAAGCAGATGAGGAGACACACAGCACTGAAT 421  
 1348 TACTTTAAGAAACACTCTCAAAACAGAAACAGACAGCAGCTGTTGAGAAATCTCG 1407  
 Db 422 TACCTGAAGAAACAAATGTTTGAAGAGCGACCTCAATTTGATGAGAGATTCATTG 481  
 1408 GAATTTATGAGCGAAGAGCTGTGTAAGCTGAGAGATATCGATCTGAGACAGAGA 1467  
 Db 482 AGTGTGTTATCCCAAGAACTTGAAGAGCAATGAGAAAGAAATCTATGAGAAATTA 541  
 1468 ACTAAGATGACAGATGAACAGAACAGGAGAGATGATGACACAGACAGTTTTTCATG 1527  
 Db 542 GCCAAGAAAGACACTTGAGATGATGAAGAGATGAAGATGACAGAGAAATTTTCCAT 601  
 1528 GATTCAATCAACCAATCTATGAAGAAAGAGCGAAAGAGAGAAATTCAGATGTG 1587  
 Db 602 GATCAAAATCGAAAAAATTCACAGGCCACAGAGGAGAAAGAAATCAAGTTTGAAGTTG 661  
 1588 CAGCAGCAGAAACGCGCAAGGTTGTGGCCAGCAGCAGCAAGCAATTAATCCCTCTAGC 1647  
 Db 662 CTCGAAGAGACCGCAGCAAGGCT-----AAGCAAGTGAATTTGACTCTGGAAT 712  
 1648 AATGACATTTGCCAAGAGAGAGCTGAGAAAGTGTCAAGCTTCATGAGTTTCAAGAGAA 1707  
 Db 713 ACTGAAGATCGAGGCAAGAGAAAGAAAGAAAGATTCAAGACTGACGAGTGAAAG 772  
 1708 GAGATGAGAGAGTTGTGTGAAGAGAGGAGATGCTGATTAAGATCAAGAGAAAGATG 1767  
 Db 773 GATGTTGAGATTTGAAGCCGAAAGCAACATGATCAATTTGATCAAGAGAAAGAG 832  
 1768 GAAGCATGAAGAAAGAGCATCAGAGAGATTTGATCTGAGAAAGAAATTTGATGAG 1827  
 Db 833 GTGAAGCTCAAGAGAGATCTGCGCAAGAAATTTGAGCTTGAGAGAGAGCTGAGACACA 892  
 1828 GCTTTGAACAGCTCATTAAGCA 1853  
 Db 893 GCCCTCACTCTCCTGATGATTAACA 918

RESULT 13  
 AF469493 513 bp mRNA linear PLN 02-JUL-2003  
 LOCUS Triticum aestivum HOTR mRNA, partial cds.  
 DEFINITION AF469493  
 ACCESSION AF469493.1 GI:32400755  
 VERSION  
 KEYWORDS Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 513)  
Li, J.R., Wang, F., Li, Q.Z. and Zhang, X.S.  
Gene isolation and expression of a new Zn-finger  
Unpublished  
2 (bases 1 to 513)  
Li, J.R., Wang, F., Li, Q.Z. and Zhang, X.S.  
Direct Submission  
Submitted (18-JUN-2002) College of Life Science, Shandong  
Agricultural University, Dai zong Street 61, Tai'an, Shandong  
271018, P.R. China

FEATURES  
source  
Location/Qualifiers  
1..513  
/organism="Triticum aestivum"  
/mol\_type="rRNA"  
/cultiyar="PH 82-2-2"  
/db\_xref="taxon:4565"  
/dev\_stage="seeds"  
/note="similar to HOF of Hordeum vulgare"  
/product="HOF"  
/protein\_id="AAP80610.1"  
/db\_xref="GI:32400756"  
/translation="MAGNOLIELEFSDVATKARHAYPGHGMVILPSSAVG  
YMEERLHMHFIDORTDRDWRVPPPGSGPOLXGLAREDEMTNRNCKSKSR  
LKYEMRSHNMVVAOMKMSDNOQLNLYKNVKTREOSKVEETLAVITCKLEETM  
ENIFVRSKA"

CDS  
10.2%; Score 192; DB 8; Length 513;  
Best Local Similarity 62.3%; Pred. No. 8.5e-36;  
Matches 319; Conservative 0; Mismatches 190; Indels 3; Gaps 1;  
Query Match  
1 TGGTTCGCGATGGGACCAAGAGCTGTGGAATCTTCAGCAAGTGTGAGGCTTTAG 1023  
1 TGGTTCGCGATGGGACCAAGAGCTGTGGAATCTTCAGCAAGTGTGAGGCTTTAG 60  
1024 GCAGCCCATCTTATGTCACAGGCGCATGTGGATGATGTTCTGATTTGAGAGC 1083  
61 GCAGCCCATCTTATGTCACAGGCGCATGTGGATGATGTTCTGATTTGAGAGC 120  
1084 AGTGCACATGCTATTGAGGCGCGAAGCGCTTCAACGGGAGTTAGCTGAGATGGGTTA 1143  
121 TCTGCTGTGGCTATATGAGGCAAGACGCTTCATAGCATTTTATTGATCAAGAACA 180  
1144 GATGAAATTCCTG--GGGTCAAGAGCGCATGTTTCTGAGAGTGTTCGCCAATG 1200  
181 GACAGGAGACCTTGGAGAGATCGAGGCTTCTTCTTACCTGTGGAGAGAGCAATTA 240  
1201 TATGGCTTCCTTCAAGAGCAAGATCTGACATATTCATCAACTCTTCAGAGCAAA 1260  
241 TACGGTTCTTCAAGAGCAAGATCTGACATATTCATCAACTCTTCAGAGCAAA 300  
1261 ACAAGGTAATTCAGATGTAATCAATCAAGAGATGTTGTAAGAGAGTGAAGCAG 1320  
301 AGCGGCTGAATACGATGATGATCTCATATGATGATGATGATGATGATGATGAT 360  
1321 ATCTCTAGAGCATGACAGCTGATGATCTTATAGAGCAAGCTTCAAAAGCAAG 1380  
361 ATAGGCAAGATATCAACATCAATCTGAGAGCAAGAGTGTGTAAGCAGAGCAG 420  
1381 CACGCCAAGTGTCTGAGAGATCTCTGAAATTTAGAGAGAGAGTGTGTAAGTGA 1440  
421 CGCTCTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
1441 GAGGTAATCGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1472  
481 GAAAGATATATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512

RESULT 14  
CN508CAO/c 91053 bp DNA linear PLN 21-NOV-2003  
LOCUS  
DEFINITION  
CN508CAO  
Oryza sativa chromosome 12, BAC OJ1561.A05 of library Monsanto  
from chromosome 12 of cultivar Nipponbare of ssp. Japonica of Oryza  
sativa (rice), complete sequence.

ACCESSION  
AL818111.4 GI:28892661  
HTG.  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriocaridaceae; Oryzaceae; Oryza.

REFERENCE  
AUTHORS  
1 (bases 1 to 91053)  
Cholsene, N., Orjeda, G., Catrolicco, L., Demange, N., Wincker, P.,  
Seguene, B., Pelletier, B., Scarpelli, C., Salanoubat, M.,  
Weissenbach, J., and Queller, P.  
Oryza sativa chromosome 12 sequencing  
Unpublished  
2 (bases 1 to 91053)  
Genoscope.  
Direct Submission  
Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage ;  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
On Mar 9, 2003 this sequence version replaced gi:28460590.  
Center: Genoscope / Centre National de Sequencage

COMMENT  
The following sequence is oriented from the T7 to the SP6 end. The  
nucleotide sequence of this BAC clone was generated by combining  
Monsanto, Syngenta and Genoscope sequencing data. (AC-BX35967)  
Upstream BAC (overlapping the T7 end) : OSJNBa0029N15 (AC-BX35967)  
Downstream BAC (overlapping the SP6 end) : OJ102\_B11 (AC-AL713901)  
----- Finishing boundaries  
FINISHED SEGMENT STARTS AT BASE 1  
FINISHED SEGMENT ENDS AT BASE 75085  
Contact: sequef@genoscope.cns.fr

FEATURES  
source  
Location/Qualifiers  
1..91053  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultiyar="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:39947"  
/chromosome="12"  
/clone\_lib="Monsanto"

ORIGIN  
Query Match  
Best Local Similarity 9.0%; Score 169.4; DB 8; Length 91053;  
Matches 310; Conservative 0; Mismatches 191; Indels 6; Gaps 2;  
Query Match  
1 TGGTTCGCGATGGGACCAAGAGCTGTGGAATCTTCAGCAAGTGTGAGGCTTTAG 519  
1 TGGTTCGCGATGGGACCAAGAGCTGTGGAATCTTCAGCAAGTGTGAGGCTTTAG 81462  
520 TTGGATGATTCGATGAGAGAGCTTCAAGATGATGATGATGATGATGATGATGAT 579  
81461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 81402  
580 AGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 639  
81401 AGTTTGAAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 81342  
640 TTGTCGATGAG 699  
81341 TTAAGTGAAG 81282

QY 700 GGACCTGTGTCATGATTGTAATA---CTGCAACCTCTCTACTAGCTCATGCGAGACA 756  
 DB 81281 GGACCTGTGTCATGATTGTAATA---CTGCAACCTCTCTACTAGCTCATGCGAGACA 81222  
 QY 757 AAAGAGCTAGGCGAGTTAAGCTCCATAGAGAAATGGCTGAAGTTTAAAGAAAGATCTA 816  
 DB 81221 AAGGGTTCTATTAAGGTCAAGGCTCACAGAAATGGCTAGTATTATGAGAAAGAGCTA 81162  
 QY 817 CAGATGAGAGGCGCATCTGTCATTCCTTGGTGAGATTATAGGCGAGTGAAGGGTTG 876  
 DB 81161 TCTCCGAGGGGAATCTTCACTGATGATCTTGGCGAACTTTAGGAAATGGAAGGATTTG 81102  
 QY 877 GGTGAGATGAAAGAGATTATGAATTTCTGCGCTCAATGCTCATCATCATGATACT 936  
 DB 81101 CG---GGAAGACACGATGATGATGATTTGTTGGCCCAATGATGATGATGATGAT 81045  
 QY 937 AGACTGATTAAGAGCATTAAGATTAAG 963  
 DB 81044 GTATTGAAACAAGATGAGATTAAG 81018

RESULT 15  
 CENS07Y01/c 112231 bp DNA linear PLN 21-NOV-2003  
 LOCUS Oryza sativa chromosome 12, . BAC OJ1102\_B11 of library Monsanto  
 DEFINITION from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza  
 sativa (rice), complete sequence.  
 AL713901 GI:38490465  
 VERSION HTG.  
 KEYWORDS Oryza sativa (japonica cultivar-group)  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 112231)  
 AUTHORS Choinsu,N., Orjeda,G., Catcolico,L., Demange,N., Wincker,P.,  
 Segurens,B., Pelletier,F., Scarpelli,C., Salanoubat,M.,  
 Weissenbach,J. and Quetier,F.  
 TITLE Oryza sativa chromosome 12 sequencing  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 112231)  
 AUTHORS Genoscope.  
 JOURNAL Direct Submission  
 TITLE Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
 COMMENT - Web : www.genoscope.cns.fr  
 On Nov 21, 2003 this sequence version replaced gi:23094326.  
 Center: Genoscope / Centre National de Sequencage  
 Center code: GS  
 Web site: http://www.genoscope.cns.fr/  
 Contact: Sequef@genoscope.cns.fr

The following sequence is oriented from the T7 to the SP6 end. The  
 nucleotide sequence of this BAC clone was generated by combining  
 Monsanto and Genoscope sequencing data.  
 Upstream BAC (overlapping the T7 end) : OJ1561\_A05 (AC-AL831811)  
 Downstream BAC (overlapping the SP6 end) : OJ1003\_A04 (AC-AL713947)  
 ----- Finishing boundaries  
 FINISHED SEGMENT STARTS AT BASE 1  
 FINISHED SEGMENT ENDS AT BASE 112231

## FEATURES

Location/Qualifiers  
 1..112231  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:39947"  
 /chromosome="12"  
 /clone="OJ1102\_B11"  
 /clone\_1fb="Monsanto"

ORIGIN

Query Match 9.0%; Score 169.4; DB 8; Length 112231;  
 Best Local Similarity 61.1%; Pred. No. 2.6e-30;  
 Matches 310; Conservative 0; Mismatches 191; Indels 6; Gaps 2;

QY 460 TTCTCTGACGTGAGAGATGATGATTAATGCTTCTGAGAAAGAGATGATTCGATGCT 519  
 DB 8800 TTGCTCTGTTGATGATTCCTGAAACAGACATGCTGAAGGTTATGACATCGGACATGAT 8741  
 QY 520 TTGATGATTTGATGAGACGCTTCGAAAGTATGATTTAGTACTCGATGATGATGAAAG 579  
 DB 8740 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8681  
 QY 580 ACCCATGATCAGCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 639  
 DB 8680 AGTTTGAAGCTCGGAAATATCAAGTTTCTTGAAGCTTCTGGAAGCTTCTGGAAGG 8621  
 QY 640 TTGTCATGACAGAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 699  
 DB 8620 TTAAGTATGAAACAGTTAAATGAACTAGGCAATGCTGCTGCTGCTGCTGCTGCTGCT 8561  
 QY 700 GGACCTGTGTCATGATTGTAATA---CTGCAACCTCTCTACTAGCTCATGCGAGACA 756  
 DB 8560 GGACCTGTGTCATGATTGTAATA---CTGCAACCTCTCTACTAGCTCATGCGAGACA 8501  
 QY 757 AAAGAGCTAGGCGAGTTAAGCTCCATAGAGAAATGGCTGAAGTTTAAAGAAAGATCTA 816  
 DB 8500 AAGGGTTCTATTAAGGTCAAGGCTCACAGAAATGGCTAGTATTATGAGAAAGAGCTA 8441  
 QY 817 CAGATGAGAGGCGCATCTGTCATTCCTTGGTGAGATTATAGGCGAGTGAAGGGTTG 876  
 DB 8440 TCTCCGAGGGGAATCTTCACTGATGATCTTGGCGAACTTTAGGAAATGGAAGGATTTG 8381  
 QY 877 GGTGAGATGAAAGAGATTATGAATTTCTGCGCTCAATGCTCATCATCATGATACT 936  
 DB 8380 CG---GGAAGACACGATGATGATGATTTGTTGGCCCAATGATGATGATGATGAT 8324  
 QY 937 AGACTGATTAAGAGCATTAAGATTAAG 963  
 DB 8323 GTATTGAAACAAGATGAGATTAAG 8297

Search completed: March 29, 2004, 23:22:40  
 Job time : 7093.65 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 08:05:52 ; Search time 4261.12 Seconds  
(without alignments)  
13161.120 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878

Sequence: 1 atgagttctagggtctgtctc.....cttcacatgaagatgatga 1878

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estbba:\*  
2: em\_estbba:\*  
3: em\_estbba:\*  
4: em\_estbba:\*  
5: em\_estbba:\*  
6: em\_estbba:\*  
7: em\_estbba:\*  
8: em\_estbba:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estfun:\*  
17: em\_gss\_num:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rdg:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	565	30.1	693	28	AQ959658 LEREM21TF
2	534	28.4	606	28	AQ959657 LEREM21TF
3	513	27.3	619	9	AV566465 AV566465
4	512	27.3	647	9	AU236368 AU236368

Result No.	Score	Query Match	Length	ID	Description
5	474	25.2	613	28	AQ964580 LEREM21TF
6	329	17.5	380	9	AV525508 AV525508
7	295	15.7	457	29	CNS0009H CNS0009H
8	273	14.5	523	28	AQ959659 LEREM21TF
9	244	13.0	421	9	AU227277 AU227277
10	228	12.1	290	28	AQ010650 AQ010650
11	169	9.0	548	9	A1999551 A1999551
12	153	8.1	197	28	BH169457 BH169457
13	148	7.9	443	28	BH618393 BH618393
14	57	3.0	482	28	BH169448 BH169448
15	47	2.5	232	28	AQ964581 AQ964581
16	38	2.0	384	29	BK653272 BK653272
17	38	2.0	437	29	BK287410 BK287410
18	29	1.5	644	14	CD835509 CD835509
19	29	1.5	697	28	BH535855 BH535855
20	28	1.5	587	13	BO507998 BO507998
21	24	1.3	123	29	BK654916 BK654916
22	24	1.3	405	29	BK287411 BK287411
23	22	1.2	270	13	BY023513 BY023513
24	22	1.2	562	12	BM262455 BM262455
25	22	1.2	621	12	BJ625937 BJ625937
26	22	1.2	675	14	CA798850 CA798850
27	22	1.2	708	14	CD815300 CD815300
28	22	1.2	896	29	CG969741 CG969741
29	22	1.2	967	9	AV254322 AV254322
30	22	1.2	2684	11	AK029504 AK029504
31	21	1.1	258	10	BB578249 BB578249
32	21	1.1	319	12	BG279367 BG279367
33	21	1.1	350	14	CF598325 CF598325
34	21	1.1	376	12	BG543396 BG543396
35	21	1.1	396	14	CB831627 CB831627
36	21	1.1	402	10	BF248666 BF248666
37	21	1.1	409	14	CF371330 CF371330
38	21	1.1	414	10	BF824285 BF824285
39	21	1.1	419	10	BF249331 BF249331
40	21	1.1	424	10	BF716900 BF716900
41	21	1.1	425	14	CF260387 CF260387
42	21	1.1	432	10	BF716619 BF716619
43	21	1.1	432	10	BF249140 BF249140
44	21	1.1	433	12	BG235108 BG235108
45	21	1.1	434	10	BF248638 BF248638

## ALIGNMENTS

RESULT 1  
LOCUS AQ959658/6  
DEFINITION LEREM21TFB LRRB Arabidopsis thaliana genomic clone LEREM21, genomic survey sequence.  
ACCESSION AQ959658  
VERSION AQ959658.1 GI:6787359  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (chale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE Buell, C.R., Lin, X., Pat, G., Barnstead, M., Bowman, C., Ullrich, T., Feldhahn, T., Liang, F., Creasy, T. and Fraser, C.M.  
Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms  
unpublished (2000)  
CONTACT: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: atetigr.org

JOURNAL COMMENT  
For additional information, see <http://www.tigr.org/tcbl/at/at.html>

Seq primer: 1F  
Class: shotgun.  
Location/Qualifiers

FEATURES  
SOURCE

1..693  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="LANDBERG ERECTA"  
/db\_xref="taxon:3702"  
/clone="LEREM21"  
/note="Torgan: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.6-0.8 Kbp before ligation."

## ORIGIN

Query Match 30.1%; Score 565; DB 28; Length 693;  
Best Local Similarity 99.8%; Pred. No. 1.6e-276;  
Matches 615; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

FEATURES  
SOURCE

1..606  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="LANDBERG ERECTA"  
/db\_xref="taxon:3702"  
/clone="LEREM21"  
/note="Torgan: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.6-0.8 Kbp before ligation."

## ORIGIN

Query Match 28.4%; Score 534; DB 28; Length 606;  
Best Local Similarity 100.0%; Pred. No. 1e-260;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES  
SOURCE

1..606  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="LANDBERG ERECTA"  
/db\_xref="taxon:3702"  
/clone="LEREM21"  
/note="Torgan: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.6-0.8 Kbp before ligation."

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
COMMENT

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 606)  
Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Uterbach, T., Feldblum, T., Liang, F., Creasy, T., and Frazer, C.M.  
Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms  
Unpublished (2000)  
Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: atc@tigr.org  
For additional information, see <http://www.tigr.org/cdb/at/ac.html>  
Seq primer: 1F  
Class: shotgun.

134 AGGTCAATTTCCAGAGAAACAAGAAACAACCACTTCTGGAAAACTTGGGTTT 193  
688 AGGTCAATTTCCAGAGAAACAAGAAACAACCACTTCTGGAAAACTTGGGTTT 629  
194 CTCAGAAATTCGAATCTCTAGAGCTTGGGGTGTCTAGAGCAAGGAGGAGGAGCAAG 253  
628 CTCAGAAATTCGAATCTCTAGAGCTTGGGGTGTCTAGAGCAAGGAGGAGGAGCAAG 569  
254 TATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGTCTCGGGCATTCAAG 313  
568 TATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGTCTCGGGCATTCAAG 509  
314 CTAAACATATCTGTCTGGGGAGCAAGCTTGAACGAAAGTATGATCACTTGGGCGAC 373  
508 CTAAACATATCTGTCTGGGGAGCAAGCTTGAACGAAAGTATGATCACTTGGGCGAC 449  
374 CCCCACTGATATCTGGCCCTCTTGGAGAGAGATGAAATTTGGCAGAGAGTT 433  
448 CCCCACTGATATCTGGCCCTCTTGGAGAGAGATGAAATTTGGCAGAGAGTT 389  
434 CTGCTCAGACACAGCTGTGTGACAGAGTTCTCTGACGTGAGAGTGTGTGATATGCTT 493  
388 CTGCTCAGACACAGCTGTGTGACAGAGTTCTCTGACGTGAGAGTGTGTGATATGCTT 329  
494 CTGAGAGAGAGATGATTCGATGCTTGTGATATCTGATGACGACCTTGGAGAGTGA 553  
328 CTGAGAGAGAGATGATTCGATGCTTGTGATATCTGATGACGACCTTGGAGAGTGA 269  
554 ATTATGATCTGGAGTGTGAGTCAAAAGAGCCATGATCAAGAAAGCAAAATAGTGTTC 613  
268 ATTATGATCTGGAGTGTGAGTCAAAAGAGCCATGATCAAGAAAGCAAAATAGTGTTC 209  
614 AAAAGTCTTTGGCAGCTTGGATGCTTGTGATGACGATGAATGAACACAGAGGC 673  
208 AAAAGTCTTTGGCAGCTTGGATGCTTGTGATGACGATGAATGAACACAGAGGC 149  
674 AGTGGCAATTTCCAGCTTGTCAAGAGCAAGCTGTGTCATCGATTAACCTGAC 733  
148 AGTGGCAATTTCCAGCTTGTCAAGAGCAAGCTGTGTCATCGATTAACCTGAC 89  
734 CTCCTAGAGCTCATGC 749  
88 CTCCTAGAGCTCATGC 73

RESULT 2  
AQ959657/c 606 bp DNA linear GSS 28-JAN-2000  
DEFINITION LEREM21TF LERE Arabidopsis thaliana genomic clone LEREM21, genomic survey sequence.  
ACCESSION AQ959657  
VERSION AQ959657.1 GI:6787358  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana

336 AGCGTTGACAGAAAGTATGATTAACAATTTGTGGCAACCCCACTGTATCTGCGCTCC 395  
486 AGCGTTGACAGAAAGTATGATTAACAATTTGTGGCAACCCCACTGTATCTGCGCTCC 427  
276 ATCCGAGAGAGTAAACGCAATGCTCGGGCACTTCAACTAAATATCTGGCGGGAG 335  
546 ATCCGAGAGAGTAAACGCAATGCTCGGGCACTTCAACTAAATATCTGGCGGGAG 487  
336 AGCGTTGACAGAAAGTATGATTAACAATTTGTGGCAACCCCACTGTATCTGCGCTCC 395  
486 AGCGTTGACAGAAAGTATGATTAACAATTTGTGGCAACCCCACTGTATCTGCGCTCC 427  
426 TTTGAGAGAGATGAATTTGACAGCAAGAGAGTTCTGCTCAGCACACAGCTGTGCA 367  
456 GGAATTTCTGACGCTGAGAGATGATGTGATTAATGCTTCTGAGGAAGAAATGATTCGA 515  
366 GGAATTTCTGACGCTGAGAGATGATGTGATTAATGCTTCTGAGGAAGAAATGATTCGA 307  
516 TGCTTTGATGATTTCTGATGACGACCTTGCAGATGATTAATGATGATGATGATGATG 575  
306 TGCTTTGATGATTTCTGATGACGACCTTGCAGATGATTAATGATGATGATGATGATG 247  
576 AAAAGCCATGATCAAGAAAGCAATTAAGTGTTCAAAAGTCTTTGGCACTTGA 635  
246 AAAAGCCATGATCAAGAAAGCAATTAAGTGTTCAAAAGTCTTTGGCACTTGA 187  
636 TAGCTTGTGATTCAGACATTAATGAACACAGAGGAGGAGGAGGAGGAGGAGGAGG 695  
186 TAGCTTGTGATTCAGACATTAATGAACACAGAGGAGGAGGAGGAGGAGGAGGAGG 127  
696 GAAAGCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 749  
126 GAAAGCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 73

RESULT 3

AV566465/c 619 bp mRNA linear EST 07-SEP-2000  
 LOCUS AV566465  
 DEFINITION Arabidopsis thaliana green siliques Columbia Arabidopsis  
 thaliana cDNA clone SQ244b06F 3', mRNA sequence.  
 ACCESSION AV566465  
 VERSION AV566465.1 GI:8737917  
 KEYWORDS Arabidopsis thaliana (thale cress)  
 SOURCE EST.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 619)  
 ASAMIZU, E., NAKAMURA, Y., SATO, S. and TABATA, S.  
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 JOURNAL DNA Res. 7 (3), 175-180 (2000)  
 MEDLINE 20363093  
 PUBMED 10907847  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yama 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
 Location/Qualifiers  
 1..619  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="SQ244b06F"  
 /tissue\_type="green siliques"  
 /clone\_lib="Arabidopsis thaliana green siliques Columbia"  
 /note="Vector: pBluescriptII SK-, Site\_1: EcoRI; Site\_2:  
 XhoI"

ORIGIN  
 Query Match 27.3%; Score 513; DB 9; Length 619;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-250;  
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1291 CAAGAGATGTTGTAAGAGAGCTGAGGAGATCTCTGAGACATCAGACGCTGAATCTAC 1350  
 DB 619 CAAGAGATGTTGTAAGAGAGCTGAGGAGATCTCTGAGACATCAGACGCTGAATCTAC 560  
 QY 1351 TTTAAGAAACAAGCTCTCAAAACAAGAACAGCAAGCCAGAGCTGTTGAGGAATCTCTGAA 1410  
 DB 559 TTTAAGAAACAAGCTCTCAAAACAAGAACAGCAAGCCAGAGCTGTTGAGGAATCTCTGAA 500  
 QY 1411 ATTATGAGGAGAAAGCTGCTGTAAGTCTGAGAGGATTAATCGATCTGTGACAGAGAACT 1470  
 DB 499 ATTATGAGGAGAAAGCTGCTGTAAGTCTGAGAGGATTAATCGATCTGTGACAGAGAACT 440  
 QY 1471 AAGATGCAAGATGAACAGAAACAGGAGAGATGATGATGACACAGACGTTTTCATGATGAT 1530  
 DB 439 AAGATGCAAGATGAACAGAAACAGGAGAGATGATGATGACACAGACGTTTTCATGATGAT 380  
 QY 1531 TCAATCAAAACAGATCCATTAAGAAAGAGCGCAAGAGAGAGATTTTCAGATGTTGCGAG 1590  
 DB 379 TCAATCAAAACAGATCCATTAAGAAAGAGCGCAAGAGAGAGATTTTCAGATGTTGCGAG 320  
 QY 1591 CAGCGAGAAAGTGCAGAGGTTGTTGGCCAGCGACGACAGACATTAATCCCTCTGACAT 1650  
 DB 319 CAGCGAGAAAGTGCAGAGGTTGTTGGCCAGCGACGACAGACATTAATCCCTCTGACAT 260  
 QY 1651 GACGATTCGCGAAGAGAGCTGAGAAAGTCAAGCTTCATGAGATTTCAAGAGAAAGAG 1710  
 DB 259 GACGATTCGCGAAGAGAGCTGAGAAAGTCAAGCTTCATGAGATTTCAAGAGAAAGAG 200  
 QY 1711 ATGAGAGAGTTTGTGAGAGAGAGGAGATGCTGATTAAGATCAAGAGAAAGATGAGAA 1770  
 DB 199 ATGAGAGAGTTTGTGAGAGAGAGGAGATGCTGATTAAGATCAAGAGAAAGATGAGAA 140

QY 1771 GACATGAAGAGAGGATCATCAGAGATATTTT 1803  
 DB 139 GACATGAAGAGAGGATCATCAGAGATATTT 107

RESULT 4  
 LOCUS AU236368  
 DEFINITION AU236368 647 bp mRNA linear EST 01-APR-2002  
 mRNA sequence.  
 ACCESSION AU236368  
 VERSION AU236368.1 GI:19875537  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 647)  
 SEKI, M., NARUSAKA, M., ICHIDA, J., KAMIYA, A., SATOU, M., NAKAJIMA, M.,  
 OONO, Y., SAKURAI, T., CARNINCI, P., KAWAI, J., ITOH, M., ISHII, Y.,  
 ARAKAWA, T., SHIBATA, K., SHINAGAWA, A., MURAMATSU, M., HAYASHIZAKI, Y.  
 and SHINOZAKI, K.  
 Large scale analysis of Arabidopsis full-length cDNA (2002b)  
 Unpublished (2002)  
 Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060  
 Email: msekic@rcc.riken.go.jp  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
 and XhoI was ligated to modified lambda PhiC-1 vector (Carninci et  
 al., submitted for publication) digested with BamHI and SalI. This  
 clone is in a modified pBluescript vector. Please visit our web  
 site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further  
 details.

ORIGIN  
 Query Match 27.3%; Score 512; DB 9; Length 647;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-249;  
 Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTTCAGAGGCTGCTCCAAATGCTTAAGAAAGAAAGCTTCAGAGCTGCTTAAGCCCT 60  
 DB 92 ATGAGTTTCAGAGGCTGCTCCAAATGCTTAAGAAAGAAAGCTTCAGAGCTGCTTAAGCCCT 151  
 QY 61 GAGTTTGAACAGTTGTTTCAAGTTTGGCAGAGAGAGAGCTGCTTCTTCAACAATGAT 120  
 DB 152 GAGTTTGAACAGTTGTTTCAAGTTTGGCAGAGAGAGAGCTGCTTCTTCAACAATGAT 211  
 QY 121 GGAGAGAGTGGAGAGCTATTTCAGAGAAAGAAAGAAAGAAAGCAAGTCTGGA 180  
 DB 212 GGAGAGAGTGGAGAGCTATTTCAGAGAAAGAAAGAAAGAAAGCAAGTCTGGA 271  
 QY 181 AAAAAGTTGTTTCTCAAGATTCGAATCTCTTAAGCTTGGGTTGTCAGAGAGAGG 240  
 DB 272 AAAAAGTTGTTTCTCAAGATTCGAATCTCTTAAGCTTGGGTTGTCAGAGAGAGG 331  
 QY 241 AGAGGTAGCAACGTATCTGGAGAGAGAAACAATGTATCCGGAGAGTAAAGCAATGGT 300



Db 332 AGAGTACAGATGATCTGGAGAGAGAAACATGATATCCGGAGAGGTAAAGGCAATGCT 391  
 Oy 301 CGGGGATTCACAGCTAATATCTGGTGGGGAGAGGCTTGAGAGAAAGATATGATAC 360  
 Db 392 CGGGGATTCACAGCTAATATCTGGTGGGGAGAGGCTTGAGAGAAAGATATGATAC 451  
 Oy 361 AACTTTGGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGATGAAATTGGCAG 420  
 Db 452 AACTTTGGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGATGAAATTGGCAG 511  
 Oy 421 GCAAGAGAGGTTCTGCTACACACAGCTTGAGAGGAGTTCCGAGCTGAGAGATGAT 480  
 Db 512 GCAAGAGAGGTTCTGCTACACACAGCTTGAGAGGAGTTCTGAGCTGAGAGATGAT 571  
 Oy 481 GTGATTAATGCTTCTGAGAGAGAGATGATTC 512  
 Db 572 GTGATTAATGCTTCTGAGAGAGAGATGATTC 603

RESULT 5  
 A0964580/c 613 bp DNA linear GSS 28-JAN-2000  
 LOCUS LERGX20TF LERG Arabidopsis thaliana genomic clone LERGX20, genomic  
 DEFINITION survey sequence.  
 ACCESSION A0964580  
 VERSION A0964580.1 GI:6792281  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana

REFERENCE  
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 613)  
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Ulterbach, T.,  
 Feldhym, T., Liang, F., Creasy, T. and Fraser, C.M.  
 TITLE Genomic survey sequencing of landsberg ecotype of  
 Arabidopsis thaliana and identification of sequence-based  
 polymorphisms  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atc@tigr.org  
 For additional information, see <http://www.tigr.org/cdb/ac/at.html>  
 Seg primer: TP  
 Class: shotgun.

## FEATURES

Source  
 1..613  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="LANDSBERG ERECTA"  
 /db\_xref="taxon:3702"  
 /clone="LERGX20"  
 /clone\_lib="LERG"  
 /note="Organ: Leaf; Vector: pUC19TK; Total genomic DNA was  
 sheared to 0.4-0.7 Kbp before ligation."

## ORIGIN

Query Match 25.2%; Score 474; DB 28; Length 613;  
 Best Local Similarity 99.8%; Pred. No. 4.5e-230;  
 Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 ATGAGTTCTAGGGCTGTGTCATATGCTTAAGAGAAAGAGCTTCAGGGTGTATAGGCT 60  
 Db 525 ATGAGTTCTAGGGCTGTGTCATATGCTTAAGAGAAAGAGCTTCAGGGTGTATAGGCT 466  
 Oy 61 GAGGTTGAACAGTTGTTCAAGGTTTGGCAGGAGCAGACTGGCTTCTTCAAGATGAT 120  
 Db 465 GAGGTTGAACAGTTGTTCAAGGTTTGGCAGGAGCAGACTGGCTTCTTCAAGATGAT 406

Oy 121 GGAGAGAGTGGAGGCTATTTCCAGAAAGAACAAACAGAAACACTTCTGCA 180  
 Db 405 GGAGAGAGTGGAGGCTATTTCCAGAAAGAACAAACAGAAACACTTCTGCA 346  
 Oy 181 AAAACTTGGGTTTCTCAGAAATTCGAATCTCTTAGAGCTTGGGGTGTGAGCAGAGG 240  
 Db 345 AAAACTTGGGTTTCTCAGAAATTCGAATCTCTTAGAGCTTGGGGTGTGAGCAGAGG 286  
 Oy 241 AGAGGTGCAACGATATCTGGAGAGAGAAACATGATCCGGAGAGGTAAAGGCAATGCT 300  
 Db 285 AGAGGTGCAACGATATCTGGAGAGAGAAACATGATCCGGAGAGGTAAAGGCAATGCT 226  
 Oy 301 CGGGGATTCACAGCTAATATCTGTGCGGAGCAGAGCTTGAGCAGAAAGTATGATAC 360  
 Db 225 CGGGGATTCACAGCTAATATCTGTGCGGAGCAGAGCTTGAGCAGAAAGTATGATAC 166  
 Oy 361 AACTTTGGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGATGAAATTGGCAG 420  
 Db 165 AACTTTGGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGATGAAATTGGCAG 106  
 Oy 421 GCAAGAGAGGTTCTGCTACACACAGCTTGAGAGAGTTTCTTGAAGAGATGAT 480  
 Db 105 GCAAGAGAGGTTCTGCTACACACAGCTTGAGAGAGTTTCTTGAAGAGATGAT 46  
 Oy 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGAT 525  
 Db 45 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGAT 1

RESULT 6  
 AV525508 380 bp mRNA linear EST 01-SEP-2000  
 LOCUS AV525508 Arabidopsis thaliana aboveground organs two to six-week  
 DEFINITION old Arabidopsis thaliana cDNA clone APD5d02R 5', mRNA sequence.  
 ACCESSION AV525508.1 GI:6685036  
 VERSION AV525508  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana

REFERENCE  
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 380)  
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 JOURNAL DNA Res. 7 (3), 175-180 (2000)  
 MEDLINE 20363093  
 PUBMED 10907847

COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

## FEATURES

Source  
 1..380  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="APD5d02R"  
 /tissue\_type="aboveground organs"  
 /dev\_stage="two to six-week old"  
 /clone\_lib="Arabidopsis thaliana aboveground organs two to  
 six-week old"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

## ORIGIN

Query Match 17.5%; Score 329; DB 9; Length 380;  
 Best Local Similarity 99.7%; Pred. No. 4e-156;  
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 276 ATCCGGAGAGGTAACGGCAATGTCGGGCAATTCAGCTAACATATCTGATCGGGAGC 335  
 Db 1 ATCCGGAGAGGTAACGGCAATGTCGGGCAATTCAGCTAACATATCTGATCGGGAGC 60  
 QY 336 AGCGTTGACAGAAAGTATGATTAACAATCTTTGGGACACCCCACTGTATCTGCCCTTC 395  
 Db 61 AGCGTTGACAGAAAGTATGATTAACAATCTTTGGGACACCCCACTGTATCTGCCCTTC 120  
 QY 396 TTGGAGAGGAGATGGAATTTGGACGACAGAGAGGTTGCTGACGACACAGCTGTGCA 455  
 Db 121 TTGGAGAGGAGATGGAATTTGGACGACAGAGAGGTTGCTGACGACACAGCTGTGCA 180  
 QY 456 GGAGTTTCTCGACGTGAGAGATGATGTGATTAATGCTTCTGAGAGAGATGATTCGA 515  
 Db 181 GGAGTTTCTCGACGTGAGAGATGATGTGATTAATGCTTCTGAGAGAGATGATTCGA 240  
 QY 516 TGCCTTGATGATTTCTGATGACGACCTTGCAAGTATGATTAATGCTGAGATGTGATCA 575  
 Db 241 TGCCTTGATGATTTCTGATGACGACCTTGCAAGTATGATTAATGCTGAGATGTGATCA 300  
 QY 576 AAAGAGCCATGATCAACGAAGCAATAGTGTTCAAAGTCTTTGGAGCTTGA 635  
 Db 301 AAAGAGCCATGATCAACGAAGCAATAGTGTTCAAAGTCTTTGGAGCTTGA 360  
 QY 636 TAGCTTGTCATCGACGACA 655  
 Db 361 TAGCTTGTCATCGACGACA 380

RESULT 7  
 CNS00P9H 457 bp DNA linear GSS 28-JUN-1999  
 LOCUS Arabidopsis thaliana genome survey sequence T7 end of BAC F821 of  
 DEFINITION IGF library from strain Columbia of Arabidopsis thaliana, genomic  
 survey sequence.  
 AL084227  
 AL084227.1 GI:5285367  
 GSS.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 457)  
 Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P.,  
 Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.

REFERENCE  
 AUTHORS  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Genoscope.  
 Direct Submision  
 Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 Location/Qualifiers  
 1..457  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="F8G21"  
 /clone\_11b="IGF"  
 /note="end : T7"

ORIGIN  
 Query Match 15.7%; Score 295; DB 29; Length 457;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-119;  
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 AGTGCCTCGGATGGAACCAAGAGCTGCTGGAATCTTGCACAAAGTATGAGGCTTTA 1021  
 Db 55 AGTGCCTCGGATGGAACCAAGAGCTGCTGGAATCTTGCACAAAGTATGAGGCTTTA 114

QY 1022 GAGCAGCCCATTCATGCTGCAAGAGCCCATGTCGGAGTAGTGTTCATGATTTGACA 1081  
 Db 115 GAGCAGCCCATTCATGCTGCAAGAGCCCATGTCGGAGTAGTGTTCATGATTTGACA 174  
 QY 1082 GAGTGCCTCGGATGGAACCAAGAGCTGCTGGAATCTTGCACAAAGTATGAGGCTTTA 1141  
 Db 175 GAGTGCCTCGGATGGAACCAAGAGCTGCTGGAATCTTGCACAAAGTATGAGGCTTTA 234  
 QY 1142 TAGATGAAATGCTCGGGGTGAGAGCCGAGATGTTTCTGAGAGTGTTCGCCAATCT 1201  
 Db 235 TAGATGAAATGCTCGGGGTGAGAGCCGAGATGTTTCTGAGAGTGTTCGCCAATCT 294  
 QY 1202 ATGCTTCTTTCGCAACGAAGCAATGTCGACATATTCATCAACTTCACAG 1256  
 Db 295 ATGCTTCTTTCGCAACGAAGCAATGTCGACATATTCATCAACTTCACAG 349

RESULT 8  
 A0959659 523 bp DNA linear GSS 28-JAN-2000  
 LOCUS LEREM21TR LERE Arabidopsis thaliana genomic clone LEREM21, genomic  
 DEFINITION survey sequence.  
 A0959659  
 A0959659.1 GI:6787360  
 GSS.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Arabidopsis thaliana (thale cress)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 523)  
 Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Uterbach,T.,  
 Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Genomic survey sequencing of landberg erecta ecotype of  
 Arabidopsis thaliana and identification of sequence-based  
 polymorphisms  
 Unpublished (2000)  
 Contact: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atc@igf.org  
 For additional information, see <http://www.tigr.org/cdb/at/at.html>  
 Seq primer: TR  
 Class: Shotgun.

FEATURES  
 source

Location/Qualifiers  
 1..523  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="LANDSBERG ERRECTA"  
 /db\_xref="taxon:3702"  
 /clone="LEREM21"  
 /clone\_11b="LERE"  
 /note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was  
 sheared to 0.6-0.8 Kbp before ligation."

ORIGIN  
 Query Match 14.5%; Score 273; DB 28; Length 523;  
 Best Local Similarity 99.3%; Pred. No. 1.8e-127;  
 Matches 423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 167 GAAACACTTCTGGAAAACTTGGGTTTCTCAGAAATTCGATTCCTTGAAGTTGGGGTG 226  
 Db 25 GAAACACTTCTGGAAAACTTGGGTTTCTCAGAAATTCGATTCCTTGAAGTTGGGGTG 84  
 QY 227 GTACGACAGAGGAGAGAGTACCAAGTATCTGGAGAGAGAAACATATATCCGGAGAG 286  
 Db 85 GTACGACAGAGGAGAGAGTACCAAGTATCTGGAGAGAGAAACATATATCCGGAGAG 144  
 QY 287 GTACGACAGAGTGTGGGGCATTCAGAGTAAATATCTGTGTGGGGAGAGAGGCTTGACA 346  
 Db 145 GTACGACAGAGTGTGGGGCATTCAGAGTAAATATCTGTGTGGGGAGAGAGGCTTGACA 204

Wed Mar 31 08:14:27 2004

us-10-030-829-2.011go.rst

Page 6

QY 347 GAAAGTATGATACAACTTTGTGCAACCCCACTGATCTGCGCCCTCTTTGGAAGAG 406  
DB 205 GAAAGTATGATACAACTTTGTGCAACCCCACTGATCTGCGCCCTCTTTGGAAGAG 264  
QY 407 GATGCAATTTGGAGGAGGAGAGGTTCTGCTCAAGCAACAGCTGTGAGAGATTTCCCG 466  
DB 265 GATGCAATTTGGAGGAGGAGAGGTTCTGCTCAAGCAACAGATGTGAGAGATTTCCCG 324  
QY 467 ACCTGAGAGATGATGATGATTAATGCTTCTGAGGAAGAGATTCGATGCTTTGGATG 526  
DB 325 ACCTGAGAGATGATGATGATTAATGCTTCTGAGGAAGAGATTCGATGCTTTGGATG 384  
QY 527 ATTCTGATGACGACCTTGGAGGATGATTAATGCTGATGATGATGATGATGATGATG 586  
DB 385 ATTCTGATGACGACCTTGGAGGATGATTAATGCTGATGATGATGATGATGATGATG 444  
QY 587 GATCAG 592  
DB 445 GATCAG 450

RESULT 9  
AU227277/c 421 bp mRNA linear EST 23-MAR-2002  
LOCUS AU227277 RAF14 Arabidopsis thaliana cDNA clone RAF14-93-K05 3'  
DEFINITION mRNA sequence.

ACCESSION AU227277 GI:19741924  
VERSION AU227277  
KEYWORDS Arabidopsis thaliana (thale cress)  
SOURCE Arabidopsis thaliana  
ORGANISM Arabidopsis thaliana (thale cress); Embryophyta; Tracheophyta;  
Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; core eudicots;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 421)  
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Setou, M., Nakajima, M.,  
Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, Y.,  
Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A.,  
Muramatsu, M., Hayashizaki, Y. and Shimozaki, K.

JOURNAL Large scale analysis of Arabidopsis full-length cDNA  
Unpublished (2002)  
COMMENT Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: meeki@rc.riken.go.jp

FEATURES  
source  
1..421  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAF14-93-K05"  
/issue\_type="root"  
/lab\_host="DH10B"  
/clone\_lib="RAF14"  
/note="Site\_1: BamHI; Site\_2: SalI"

#### ORIGIN

Query Match 13.0%; Score 244; DB 9; Length 421;  
Best local similarity 100.0%; Pred. No. 1e-112;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1635 TAATCCCTTAGCAATGACGATTCGCAAGAGAGAGCTGAGAGAGTGTCAAGCTTCATCGA 1694

DB 421 TAATCCCTTAGCAATGACGATTCGCAAGAGAGAGCTGAGAGAGTGTCAAGCTTCATCGA 362  
QY 1695 GTTTCAGAGAGAGAGAGAGAGAGTGTGTAAGAGAGAGAGTGTGTAAGAGATCA 1754  
DB 361 GTTTCAGAGAGAGAGAGAGAGAGTGTGTAAGAGAGAGAGTGTGTAAGAGATCA 302  
QY 1755 AG 1814  
DB 301 AG 242  
QY 1815 AGAATTTGATGAGCTTTGGAG 1874  
DB 241 AGAATTTGATGAGCTTTGGAG 182  
QY 1875 TTGA 1878  
DB 181 TTGA 178

RESULT 10  
AQ010650 290 bp DNA linear GSS 29-MAY-1998  
LOCUS F27C8TRC IGF Arabidopsis thaliana genomic clone F27C8, genomic  
DEFINITION survey sequence.  
ACCESSION AQ010650 GI:3165927  
VERSION AQ010650  
KEYWORDS GSS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Arabidopsis thaliana (thale cress); Embryophyta; Tracheophyta;  
Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; core eudicots;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 290)  
Rounsley, S.D., Suh, E.J., Wible, C., Golden, K., Shatsman, S., Choi, P.,  
Yu, K., Aktinetye, B., Shen, K., Gomansekaram, S., Millican, D.,  
Adams, M.D. and Venter, J.C.  
A BAC End Sequence Database for Identifying Minimal Overlaps in  
Arabidopsis Genomic Sequencing. Update 4  
Unpublished (1998)  
JOURNAL Other GSSs: F27C8TRC  
COMMENT Contact: Steve Rounsley  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: rounsley@igr.org  
Seq primer: M13 Reverse  
Class: BAC ends

High quality sequence stop: 290.

FEATURES  
source  
1..290  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="F27C8"  
/sex="hermaphrodite"  
/clone\_lib="IGF"  
/note="Vector: Belobacti; Site\_1: EcoRI; Site\_2: EcoRI;  
Produced by Thomas Altmann"

#### ORIGIN

Query Match 12.1%; Score 228; DB 28; Length 290;  
Best local similarity 100.0%; Pred. No. 1.3e-104; Indels 0; Gaps 0;  
Matches 228; Conservative 0; Mismatches 0;

QY 962 AGTGGCTGCGATGAGGAGACCAAGAGCTGCTGATTAATTCGACAGATGAGGCTCTTA 1021

DB 63 AGTGGCTGCGATGAGGAGACCAAGAGCTGCTGATTAATTCGACAGATGAGGCTCTTA 122

QY 1022 GAGCAGCGCATTCCTATGATGTCACAGGCGCATGCTGGAGATGATGTTCTTGAGA 1081

Db 123 GAGAGCGCCATTCATATGCTCCACAGGGGCATCGGGATGATGTTCTTGATGTTTGACA 182  
OY 1082 GGAGTGGCCACTGCTATTTGAGAGCGGAGCTCCACCGGGAGTTAGTGAATGGAGT 1141  
Db 183 GAGAGTGGCCACTGCTATTTGAGAGCGGAGCTCCACCGGGAGTTAGTGAATGGAGT 242  
OY 1142 TAGATAGATTCCTGGGGTCAAGAGCGCATATGTTTCTGAGGTG 1189  
Db 243 TAGATAGATTCCTGGGGTCAAGAGCGCATATGTTTCTGAGGTG 290  
RESULT 11 548 bp mRNA linear EST 08-SEP-1999  
LOCUS A1999551.1  
DEFINITION thaliana CDNA clone 701556368, mRNA sequence.  
ACCESSION A1999551  
VERSION A1999551.1 GI:5846456  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 548)  
AUTHORS Chen, J., Momiyama, M., Chan, E., Mooney, M., Garroun, B., Gilliland, D.,  
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brooker, P.,  
Gorgone, G., Burns, D., Griffin, J., Mouanoudou, M., Nguyen, D.,  
Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Bottino, C.,  
Cardio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S.,  
Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L., and  
Hanson, D.  
TITLE Arabidopsis thaliana Gene Expression Microarray  
JOURNAL Unpublished (1999)  
COMMENT Contact: David Smolter, Ph.D.  
Genome Systems, Inc., a wholly owned subsidiary of Incyte  
Pharmaceuticals, Inc.  
4633 World Parkway Circle, St. Louis, MO 63134, USA  
Tel: 877-577-2733  
Fax: 314-427-3324  
Email: service@genomesystems.com.  
FEATURES  
source location/Qualifiers  
1..548  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/cultivar="Columbia Col-0"  
/db\_xref="taxon:3702"  
/clone="701556368"  
/tissue\_type="rosette"  
/dev\_stage="4 - 7 weeks"  
/clone\_lib="A. thaliana, Columbia Col-0, rosette-3"  
/note="Vector: pSPORT Site\_1: NotI; Site\_2: SalI; CDNA  
library was derived from untreated rosette tissue from  
Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks.  
Plants were grown in 1:1:1 peat moss/vermiculite/perlite  
soil at 22 deg. C +/- 3 deg. C under constant light, and  
watered with fertilizer. cDNA synthesis was initiated  
using a NotI-oligo(dT) primer. Double-stranded cDNA was  
blunted, ligated to SalI adaptor, digested with NotI,  
size-selected, and cloned into the NotI and SalI sites of  
the pSPORT vector."

## ORIGIN

Query Match 9.0%; Score 169; DB 9; Length 548;  
Best Local Similarity 100.0%; Pred. No. 2.2e-74;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1710 GATGAGAGATTGTTGGAAGAGAGAGTGTCTAAAGATCAAGAGAAAGATGA 1769  
Db 345 GATGAGAGATTGTTGGAAGAGAGAGTGTCTAAAGATCAAGAGAAAGATGA 286  
OY 1770 AGACATGAAGAGAGCATCAAGAGAGATATTTGATCTGAGAAAGATTTGATGAGC 1829

Db 285 AGACATGAAGAGAGCATCAAGAGAGATATTTGATCTGAGAAAGATTTGATGAGC 226  
OY 1830 TTGGAACAGCTCATGTCAAGCATGGCTTCAATGAAGATGATGA 1878  
Db 225 TTGGAACAGCTCATGTCAAGCATGGCTTCAATGAAGATGATGA 177

RESULT 12 197 bp DNA linear GSS 03-OCT-2001  
LOCUS BH169457  
DEFINITION SALK\_001394 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
thaliana genomic clone SALK\_001394, genomic survey sequence.  
ACCESSION BH169457  
VERSION BH169457.1 GI:15904832  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 197)  
AUTHORS Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
Shinn, P., Zimmerman, J., and Boker, J.R.  
TITLE A Sequence-Indexed Library of Insertion Mutations in the  
Arabidopsis Genome  
JOURNAL Unpublished (2001)  
COMMENT Contact: Joseph R. Boker  
Salk Institute Genomic Analysis Laboratory (SIGAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: eckere@salk.edu  
This is single pass sequence recovered from the left border of  
TDNA.  
FEATURES  
source location/Qualifiers  
1..197  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="SALK\_001394"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

## FEATURES

source

Query Match 8.1%; Score 153; DB 28; Length 197;  
Best Local Similarity 100.0%; Pred. No. 2.3e-66;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## ORIGIN

OY 50 GTTATAGGCTGAGGTGAACAGTTGTTCAAGTTTGCGAGGAGCAGACTGGCTTCT 109  
Db 1 GTTATAGGCTGAGGTGAACAGTTGTTCAAGTTTGCGAGGAGCAGACTGGCTTCT 60  
OY 110 CACAAGATGATGAGAGAGTGGAGGTCATTTCCAAAGAACAGAACAAACAGGAA 169  
Db 61 CACAAGATGATGAGAGAGTGGAGGTCATTTCCAAAGAACAGAACAAACAGGAA 120  
OY 170 ACACTTGTGAAAAAATTGGTTTCTCAGAAATT 202  
Db 121 ACACTTGTGAAAAAATTGGTTTCTCAGAAATT 153

RESULT 13  
BH169457/C

LOCUS BH618393 443 bp DNA linear GSS 30-JAN-2002  
 DEFINITION SALK\_039005 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK\_039005, genomic survey sequence.  
 ACCESSION BH618393  
 VERSION BH618393.1 GI:18428488  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 443)  
 Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shin,P., Zimmerman,J. and Ecker,J.R.  
 TITLE A Sequence-indexed library of Insertion Mutations in the Arabidopsis Genome  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@salk.edu  
 This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated intron of At5g23570.  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1..443  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_039005"  
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

ORIGIN  
 Query Match 7.9%; Score 148; DB 28; Length 443;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-63;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 962 AGTGGCTGGCATGGCAACCAAGAGCTGCTGATCTGACAAAGTATGAGGCTCTTA 1021  
 |||||  
 Db 407 AGTGGCTGGCATGGCAACCAAGAGCTGCTGATCTGACAAAGTATGAGGCTCTTA 348  
 Oy 1022 GACACAGCCATCTCTATGATGTCACAGAGGCATCTGATGATGATGATGATGATGAT 1081  
 |||||  
 Db 347 GAGACGCCATCTCTATGATGTCACAGAGGCATCTGATGATGATGATGATGATGAT 288  
 Oy 1082 GCAGTGGCACTGCTATTTGAGAGCCGA 1109  
 |||||  
 Db 287 GCAGTGGCACTGCTATTTGAGAGCCGA 260

RESULT 14  
 BH169448 482 bp DNA linear GSS 03-OCT-2001  
 LOCUS BH169448  
 DEFINITION SALK\_001377 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK\_001377, genomic survey sequence.  
 ACCESSION BH169448  
 VERSION BH169448.1 GI:15904823  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE  
 1 (bases 1 to 482)  
 Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shin,P., Zimmerman,J. and Ecker,J.R.  
 TITLE A Sequence-indexed library of Insertion Mutations in the Arabidopsis Genome  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@salk.edu  
 This is single pass sequence recovered from the left border of TDNA.  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1..482  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_001377"  
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

## ORIGIN

Query Match 3.0%; Score 57; DB 28; Length 482;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-17;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 146 AGAAGACAGACAAACCAAGAGAGCTGCTGATGATGATGATGATGATGATGAT 202  
 |||||  
 Db 111 AGAAGACAGACAAACCAAGAGAGCTGCTGATGATGATGATGATGATGATGAT 167

## ORIGIN

RESULT 15  
 AO964581 232 bp DNA linear GSS 28-JAN-2000  
 LOCUS AO964581  
 DEFINITION LERGX20TR LERG Arabidopsis thaliana genomic clone LERGX20, genomic survey sequence.  
 ACCESSION AO964581  
 VERSION AO964581.1 GI:6792282  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 232)  
 Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Uterbach,T., Feldblyum,T., Liang,F., Cressy,T. and Fraser,C.M.  
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atc@tigr.org  
 For additional information, see <http://www.tigr.org/tcdb/at/at.html>  
 Seq primer: TR  
 Class: shotgun.  
 Location/Qualifiers

```
source          1. .232
                 /organism="Arabidopsis thaliana"
                 /mol_type="genomic DNA"
                 /strain="LANDSBERG ERRECTA"
                 /db_xref="taxon:3702"
                 /clone="LBRGX20"
                 /clone_11b="LBRGX"
                 /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
sheared to 0.4-0.7 Kbp before ligation."
```

## ORIGIN

```
Query Match      2.5%; Score 47; DB 28; Length 232;
Best Local Similarity 100.0%; Pred. No. 3.2e-12;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      6 TTCTAGGGGCTGCTCCAAATGCTTAAGGAAAGAAACGTTCAAGGTCGTT 52
          |||||||
Db      134 TTCTAGGGGCTGCTCCAAATGCTTAAGGAAAGAAACGTTCAAGGTCGTT 180
```

Search completed: March 30, 2004, 18:05:22  
Job time : 4271.12 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 08:22:12 ; Search time 140.312 Seconds  
(without alignments)  
7427.693 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878  
Sequence: 1 atgagcttcaggctgctgc.....ttcacatgaagatgatga 1878

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents, NA: \*  
1: /cgn2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/ECTUS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.0	481	4	US-09-615-192A-109 Sequence 109, App
C 2	19	1.0	481	4	US-09-169-789-109 Sequence 109, App
C 3	19	1.0	495	2	US-08-975-316-27 Sequence 27, Appl
C 4	19	1.0	495	4	US-09-615-192A-27 Sequence 27, Appl
C 5	19	1.0	495	4	US-09-169-789-27 Sequence 27, Appl
6	19	1.0	2000	1	US-09-041-075A-4 Sequence 4, Appl
7	19	1.0	2000	1	US-09-041-075A-6 Sequence 6, Appl
8	19	1.0	8257	4	US-09-595-684B-30 Sequence 30, Appl
9	19	1.0	8503	4	US-09-620-312D-130 Sequence 130, App
10	19	1.0	1664976	4	US-08-916-421B-1 Sequence 1, Appl
11	18	1.0	462	4	US-09-489-039A-5315 Sequence 5315, Ap
12	18	1.0	1254	4	US-09-322-478-25 Sequence 25, Appl
13	18	1.0	1829	1	US-07-688-352C-15 Sequence 15, Appl
14	18	1.0	1829	3	US-08-146-379C-15 Sequence 15, Appl
15	18	1.0	1829	3	US-09-146-249A-15 Sequence 15, Appl
16	18	1.0	1829	3	US-08-206-188B-15 Sequence 15, Appl
17	18	1.0	1829	5	PCT-US91-02714-15 Sequence 15, Appl
18	18	1.0	1952	4	US-09-533-028-89 Sequence 89, Appl
C 19	18	1.0	2001	4	US-09-489-039A-5217 Sequence 5217, Ap
C 20	18	1.0	3031	1	US-08-785-241-2 Sequence 2, Appl
C 21	18	1.0	7478	4	US-10-104-966-15 Sequence 15, Appl
22	18	1.0	8906	2	US-08-826-267-1 Sequence 1, Appl
23	18	1.0	35524	3	US-08-923-137-1 Sequence 3, Appl
24	18	1.0	42571	4	US-09-810-347-3 Sequence 3, Appl
25	18	1.0	100848	4	US-09-596-002-39 Sequence 39, Appl
26	17	0.9	20	4	US-09-851-896-62 Sequence 62, Appl
C 27	17	0.9	47	4	US-09-422-978-3322 Sequence 3322, Ap

28	17	0.9	247	4	US-09-411-977-5 Sequence 5, Appl
C 29	17	0.9	396	4	US-09-107-532A-767 Sequence 767, App
C 30	17	0.9	428	4	US-09-702-705-1214 Sequence 1214, Ap
C 31	17	0.9	428	4	US-09-736-457-1214 Sequence 1214, Ap
C 32	17	0.9	428	4	US-09-614-124B-1214 Sequence 1214, Ap
C 33	17	0.9	428	4	US-09-671-325-1214 Sequence 1214, Ap
34	17	0.9	477	4	US-09-621-976-10972 Sequence 10972, A
35	17	0.9	536	4	US-09-621-976-16220 Sequence 16220, A
36	17	0.9	561	1	US-08-565-386-3 Sequence 3, Appl
37	17	0.9	576	4	US-09-107-532A-1728 Sequence 1728, Ap
C 38	17	0.9	793	4	US-09-221-017B-940 Sequence 940, App
C 39	17	0.9	1250	3	US-09-018-584A-36 Sequence 36, Appl
C 40	17	0.9	1284	4	US-09-328-352-594 Sequence 594, Appl
C 41	17	0.9	1290	4	US-09-543-681A-1061 Sequence 1061, Ap
C 42	17	0.9	1323	4	US-09-489-039A-4961 Sequence 4961, Ap
C 43	17	0.9	1401	4	US-09-134-000C-2355 Sequence 2355, Ap
C 44	17	0.9	1428	2	US-08-960-022-7 Sequence 7, Appl
45	17	0.9	1496	4	US-09-016-434-1127 Sequence 1127, Ap

#### ALIGNMENTS

```
RESULT 1
US-09-615-192A-109/c
; Sequence 109, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-615-192A-109

Query Match          1.0% Score 19; DB 4; Length 481,
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      74 TGGTCAAGGTTTGCAGG 92
Db      470 TGGTCAAGGTTTGCAGG 452

RESULT 2
US-09-169-789-109/c
; Sequence 109, Application US/09169789
; Patent No. 6653528
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c2
; CURRENT APPLICATION NUMBER: US/09/169,789
; CURRENT FILING DATE: 1998-10-09
; EARLIER APPLICATION NUMBER: US 08/975,316
; EARLIER FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: US 08/713,000
```

EARLIER FILING DATE: 1996-09-11  
NUMBER OF SEQ ID NOS: 185  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 109  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-169-789-109

Query Match 1.0%; Score 19; DB 4; Length 481;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92  
DB 470 TGGTCAAGGTTGGCAGG 452

RESULT 3  
US-08-975-316-27/C

Sequence 27, Application US/08975316  
Patent No. 5952486  
GENERAL INFORMATION:  
APPLICANT: BLOKBERG, Leonard N., HAVUKKALA, Ilkka  
APPLICANT: and GRIERSON, Alastair W.  
TITLE OF INVENTION: MATERIALS AND METHODS FOR  
TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,316  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/713,000  
FILING DATE: September 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: SLEATH, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000/1003C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 495 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-975-316-27

Query Match 1.0%; Score 19; DB 2; Length 495;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92  
DB 484 TGGTCAAGGTTGGCAGG 466

RESULT 4

US-09-615-192A-27/C  
Sequence 27, Application US/09615192A  
Patent No. 6410718  
GENERAL INFORMATION:  
APPLICANT: Blokberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Materials and Methods for the  
TITLE OF INVENTION: Modification of Plant Lignin Content  
FILE REFERENCE: 11000.1003C4U  
CURRENT APPLICATION NUMBER: US/09/615,192A  
CURRENT FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: US 08/713,000  
PRIOR FILING DATE: 1996-09-11  
PRIOR APPLICATION NUMBER: US 09/169,789  
PRIOR FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 405  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 27  
LENGTH: 495  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-615-192A-27

Query Match 1.0%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92  
DB 484 TGGTCAAGGTTGGCAGG 466

RESULT 5  
US-09-169-789-27/C  
Sequence 27, Application US/09169789  
Patent No. 6653528  
GENERAL INFORMATION:  
APPLICANT: Blokberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Materials and Methods for the  
TITLE OF INVENTION: Modification of Plant Lignin Content  
FILE REFERENCE: 11000.1003C2  
CURRENT APPLICATION NUMBER: US/09/169,789  
CURRENT FILING DATE: 1998-10-09  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1997-11-21  
EARLIER APPLICATION NUMBER: US 08/713,000  
EARLIER FILING DATE: 1996-09-11  
NUMBER OF SEQ ID NOS: 185  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 27  
LENGTH: 495  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-169-789-27

Query Match 1.0%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92  
DB 484 TGGTCAAGGTTGGCAGG 466

RESULT 6  
US-09-041-075A-4  
Sequence 4, Application US/09041075A  
Patent No. H002022  
GENERAL INFORMATION:  
APPLICANT: Heidler, Steven A

```

; APPLICANT: Radding, Jeffrey A
; TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
; FILE REFERENCE: X-11242 Sequence Lat
; Patent No. H002022
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Candida kruisii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (300)..(1739)
US-09-041-075A-4
```

```

Query Match          1.0%; Score 19; DB 1; Length 2000;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      523 GATGATTCTGATGACGACC 541
      |||||
DB      1560 GATGATTCTGATGACGACC 1578
```

```

RESULT 7
US-09-041-075A-6
; Sequence 6, Application US/09041075A
; Patent No. H002022
; GENERAL INFORMATION:
; APPLICANT: Heidler, Steven A
; APPLICANT: Radding, Jeffrey A
; TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
; FILE REFERENCE: X-11242 Sequence Lat
; Patent No. H002022
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2000
; TYPE: RNA
; ORGANISM: Candida kruisii
US-09-041-075A-6
```

```

Query Match          1.0%; Score 19; DB 1; Length 2000;
Best Local Similarity 73.7%; Pred. No. 12;
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      523 GATGATTCTGATGACGACC 541
      |||||
DB      1560 GAUGAUCUGAUGACGACC 1578
```

```

RESULT 8
US-09-595-684B-30
; Sequence 30, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugenii
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
```

```

; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 8257
; TYPE: DNA
; ORGANISM: Human
US-09-595-684B-30
```

```

Query Match          1.0%; Score 19; DB 4; Length 8257;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1802 TTGATCTGGAGAAAGATT 1820
      |||||
DB      2141 TTGATCTGGAGAAAGATT 2159
```

```

RESULT 9
US-09-620-312D-130
; Sequence 130, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; TITLE OF INVENTION: polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/486,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_FL_genes Version 1.0
; SEQ ID NO 130
; LENGTH: 8503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (91)..(8082)
US-09-620-312D-130
```

```

Query Match          1.0%; Score 19; DB 4; Length 8503;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1802 TTGATCTGGAGAAAGATT 1820
      |||||
DB      2141 TTGATCTGGAGAAAGATT 2159
```

```
RESULT 10
```

US-08-916-421B-1  
Sequence 1, Application US/08916421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Built et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
Patent No. 6503729  
TITLE OF INVENTION: jannaschi  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA  
ORGANISM: Methanococcus jannaschi  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98343)..(98343)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (103998)..(103998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (231980)..(231980)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (23187)..(23187)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234220)..(234220)

OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (319226)..(319226)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559167)..(559167)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559241)..(559241)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (600992)..(600992)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (622708)..(622708)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657081)..(657081)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657203)..(657203)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (674435)..(674435)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (682442)..(682442)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (713552)..(713552)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (741684)..(741684)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779455)..(779455)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779676)..(779676)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature  
LOCATION: (1310988) .. (1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1313224) .. (1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349473) .. (1349473)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349491) .. (1349491)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1470091) .. (1470091)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1569020) .. (1569020)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1602912) .. (1602912)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1603734) .. (1603734)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1637998) .. (1637998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1664854) .. (1664854)  
OTHER INFORMATION: n equals a, t, c, or g  
US-08-916-421B-1

Query Match 1.0%; Score 19; DB 4; Length 1664976;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1702 GAGAAAGATGAGAGACT 1720  
DB 358045 GAGAAAGATGAGAGACT 358063

RESULT 11  
US-09-489-039A-5315  
Sequence 5315, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709 2004001  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 5315  
LENGTH: 462  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-5315

Query Match 1.0%; Score 18; DB 4; Length 462;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1616 GCCAGCAGCAGCAACA 1633  
DB 216 GCCAGCAGCAGCAACA 233

RESULT 12  
US-09-322-478-25/C  
Sequence 25, Application US/09322478

Patent No. 6331662  
GENERAL INFORMATION:  
APPLICANT: Wright, David A.  
APPLICANT: Voytas, Daniel F.  
TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto  
FILE REFERENCE: P-1065 ISURF Plant Retroelement  
CURRENT APPLICATION NUMBER: US/09/322,478  
CURRENT FILING DATE: 1999-05-28  
EARLIER APPLICATION NUMBER: 60/087125  
EARLIER FILING DATE: 1998-05-29  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 25  
LENGTH: 1254  
TYPE: DNA  
ORGANISM: Pisum sativum  
US-09-322-478-25

Query Match 1.0%; Score 18; DB 4; Length 1254;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1335 TCAGCAGCTGAACACTT 1352  
DB 890 TCAGCAGCTGAACACTT 873

RESULT 13  
US-07-688-352C-15  
Sequence 15, Application US/07688352C  
Patent No. 5527896  
GENERAL INFORMATION:  
APPLICANT: Wigler, Michael H.  
APPLICANT: Colicelli, John J.  
TITLE OF INVENTION: Cloning By Complementation and Related  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray &  
STREET: Two First National Plaza, 20 South Clark  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/688,352C  
FILING DATE: 19910419  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/511,715  
FILING DATE: 20-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Borun, Michael F.  
REGISTRATION NUMBER: 25447  
REFERENCE/DOCKET NUMBER: 27805/30197  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 346-5750  
TELEFAX: (312) 984-9740  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1829 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1421  
US-07-688-352C-15

Query Match 1.0%; Score 18; DB 1; Length 1829;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 ATTCCAGAGAGACAG 156  
DB 985 ATTCCAGAGAGACAG 1002

RESULT 14  
US-08-474-379C-15  
Sequence 15, Application US/08474379C  
Patent No. 5977305  
GENERAL INFORMATION:  
APPLICANT: Wigler, Michael H.  
COLLIGAND: Colicelli, John J.  
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED  
PROCESSES  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,379C  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/511,715  
FILING DATE: 20-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/206,188  
FILING DATE: 01-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/588,352  
FILING DATE: 19-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 27866/32771  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1829 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1421  
US-08-474-379C-15

Query Match 1.0%; Score 18; DB 2; Length 1829;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 ATTCCAGAGAGACAG 156

DB 985 ATTCCAGAGAGACAG 1002

RESULT 15  
US-09-146-249A-15  
Sequence 15, Application US/09146249A  
Patent No. 6069240  
GENERAL INFORMATION:  
APPLICANT: Wigler, Michael H.  
COLLIGAND: Colicelli, John J.  
TITLE OF INVENTION: Cloning by Complementation and Related  
Processes  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/146,249A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/511,715  
FILING DATE: 20-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1829 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1421  
US-09-146-249A-15

Query Match 1.0%; Score 18; DB 3; Length 1829;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 ATTCCAGAGAGACAG 156  
DB 985 ATTCCAGAGAGACAG 1002

Search completed: March 30, 2004, 18:11:52  
Job time : 149.312 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 14:48:39 ; Search time 1169.15 Seconds

(without alignments)  
5980.258 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878

Sequence: 1 atgagctcagcgtcgtcc.....ttcacatgaagatgatga 1878

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2458946 seqs, 1861504846 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PC7\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PC7S\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	1.1	1377	US-10-424-599-91177	Sequence 91177, A
2	20	1.1	1589	US-10-424-599-91176	Sequence 91176, A
3	19	1.0	172	US-10-085-783A-19246	Sequence 19246, A
4	19	1.0	172	US-10-242-535A-19246	Sequence 19246, A
5	19	1.0	397	US-10-424-599-57837	Sequence 57837, A
6	19	1.0	429	US-10-027-632-293476	Sequence 293476, A
7	19	1.0	481	US-10-174-693-109	Sequence 109, App1
8	19	1.0	495	US-10-174-693-27	Sequence 27, App1
9	19	1.0	501	US-09-833-790-167	Sequence 167, App1
10	19	1.0	681	US-10-424-599-18549	Sequence 18549, A
11	19	1.0	696	US-10-027-632-23555	Sequence 23555, A
12	19	1.0	696	US-10-027-632-23556	Sequence 23556, A
13	19	1.0	824	US-10-027-632-11811	Sequence 11811, A
14	19	1.0	824	US-10-027-632-11812	Sequence 11812, A
15	19	1.0	1137	US-10-369-493-24736	Sequence 24736, A

16	19	1.0	1341	9	US-09-938-842A-2465	Sequence 2465, Ap
17	19	1.0	1341	11	US-09-938-842A-2465	Sequence 2465, Ap
18	19	1.0	1576	12	US-10-424-599-4367	Sequence 4367, Ap
19	19	1.0	2000	9	US-09-742-582-4	Sequence 4, Appl1
20	19	1.0	2000	9	US-09-742-582-6	Sequence 6, Appl1
21	19	1.0	2000	10	US-09-742-580-4	Sequence 4, Appl1
22	19	1.0	2000	10	US-09-742-580-6	Sequence 6, Appl1
23	19	1.0	2000	10	US-09-742-581-4	Sequence 4, Appl1
24	19	1.0	2000	10	US-09-742-581-6	Sequence 6, Appl1
25	19	1.0	2240	15	US-10-108-260A-1451	Sequence 1451, Ap
26	19	1.0	3579	14	US-10-176-847-9	Sequence 9, Appl1
27	19	1.0	6930	10	US-09-770-107-1	Sequence 1, Appl1
28	19	1.0	7992	10	US-09-893-519A-140	Sequence 140, App
29	19	1.0	8491	14	US-10-133-013-260	Sequence 260, App
30	19	1.0	8493	13	US-10-071-766-51	Sequence 51, Appl1
31	19	1.0	8503	14	US-10-037-270-130	Sequence 130, App
32	19	1.0	8503	15	US-10-117-722-130	Sequence 130, App
33	19	1.0	47448	15	US-10-085-117-145	Sequence 145, App
34	19	1.0	1691139	14	US-10-067-514-1	Sequence 1, Appl1
35	19	1.0	1691139	15	US-10-419-723-1	Sequence 1, Appl1
36	19	1.0	2940917	15	US-10-027-632-174763	Sequence 174763, A
37	18	1.0	182	14	US-10-029-386-26108	Sequence 26108, A
38	18	1.0	189	9	US-09-864-761-32008	Sequence 32008, A
39	18	1.0	209	9	US-09-960-352-12756	Sequence 12756, A
40	18	1.0	256	12	US-10-424-599-116017	Sequence 116017, A
41	18	1.0	270	9	US-09-960-352-14374	Sequence 14374, A
42	18	1.0	304	12	US-10-424-599-112786	Sequence 112786, A
43	18	1.0	312	9	US-09-728-445-2930	Sequence 290, App
44	18	1.0	315	9	US-09-960-352-12961	Sequence 12961, A
45	18	1.0	395	9	US-09-960-352-3954	Sequence 3954, Ap

## ALIGNMENTS

RESULT 1  
US-10-424-599-91177  
; Sequence 91177, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 91177  
; LENGTH: 1377  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53343C.1  
US-10-424-599-91177

Query Match 1.1%; Score 20; DB 12; Length 1377;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 673 CAGTGGCATGTGTCAGCTTG 692  
DB 359 CAGTGGCATGTGTCAGCTTG 378

RESULT 2  
US-10-424-599-91176  
; Sequence 91176, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J



```
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 91176
LENGTH: 1589
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
US-10-424-599-91176
```

```
Query Match      1.0%; Score 19; DB 12; Length 1589;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      673 CAGTGGCATTGTCACGCTTG 692
      |||
Db      359 CAGTGGCATTGTCACGCTTG 378
```

```
RESULT 3
US-10-085-783A-19246
Sequence 19246, Application US/10085783A
Publication No. US20040037841A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 19246
LENGTH: 172
TYPE: DNA
ORGANISM: Human
US-10-085-783A-19246
```

```
Query Match      1.0%; Score 19; DB 12; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1767 GGAGACATGAGAGAGG 1785
      |||
Db      86 GGAGACATGAGAGAGG 104
```

```
RESULT 4
US-10-242-535A-19246
Sequence 19246, Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
```

```
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 19246
LENGTH: 172
TYPE: DNA
ORGANISM: Human
US-10-242-535A-19246
```

```
Query Match      1.0%; Score 19; DB 15; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1767 GGAGACATGAGAGAGG 1785
      |||
Db      86 GGAGACATGAGAGAGG 104
```

```
RESULT 5
US-10-424-599-57837
Sequence 57837, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 57837
LENGTH: 397
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_23239C.1
US-10-424-599-57837
```

```
Query Match      1.0%; Score 19; DB 12; Length 397;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1700 AAGAGAAAGATGAGGA 1718
      |||
Db      262 AAGAGAAAGATGAGGA 280
```

```
RESULT 6
US-10-027-632-293476
Sequence 293476, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
```

```
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 325720
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 293476
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-293476
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 15; Length 429;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 488 ATGCTTCTGAGAGAGAA 506
DB 76 ATGCTTCTGAGAGAGAA 94
```

```
RESULT 7
US-10-174-693-109/c
; Sequence 109, Application US/10174693
```

```
; Publication No. US2003013173A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003C5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-174-693-109
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 14; Length 481;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 74 TGGTCAAGGTTGGCAG 92
DB 470 TGGTCAAGGTTGGCAG 452
```

```
RESULT 8
US-10-174-693-27/c
```

```
; Sequence 27, Application US/10174693
; Publication No. US2003013173A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003C5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
```

```
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-174-693-27
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 14; Length 495;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 74 TGGTCAAGGTTGGCAG 92
DB 484 TGGTCAAGGTTGGCAG 466
```

```
RESULT 9
US-09-833-790-167
```

```
; Sequence 167, Application US/09833790
; Patent No. US2002068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tonglong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Indrias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-833-790-167
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 9; Length 501;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1802 TTGATCTGAGAGAGATT 1820
DB 10 TTGATCTGAGAGAGATT 28
```

```
RESULT 10
US-10-424-599-18549/c
; Sequence 18549, Application US/10424599
; Publication No. US20040031072A1
```

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
```

```
SEQ ID NO 18549
LENGTH: 681
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_116754C.1
US-10-424-599-18549
```

```
Query Match      1.0%; Score 19; DB 12; Length 681;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1745 TAAAGATCAGAGAA 1763
DB      371 TAAAGATCAGAGAA 353
```

```
RESULT 11
US-10-027-632-23555/c
Sequence 23555, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
```

```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23555
LENGTH: 696
TYPE: DNA
ORGANISM: Human
US-10-027-632-23555
```

```
Query Match      1.0%; Score 19; DB 15; Length 696;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1122 GGAGTTAGCTGAGTGGG 1140
DB      82 GGAGTTAGCTGAGTGGG 64
```

```
RESULT 12
US-10-027-632-23556/c
Sequence 23556, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
```

```
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23556
LENGTH: 696
TYPE: DNA
ORGANISM: Human
US-10-027-632-23556
```

```
QY      1122 GGAGTTAGCTGAGTGGG 1140
DB      82 GGAGTTAGCTGAGTGGG 64
```

```
RESULT 13
US-10-027-632-11811
Sequence 11811, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
```

```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11811
LENGTH: 824
TYPE: DNA
ORGANISM: Human
US-10-027-632-11811
```

```
Query Match      1.0%; Score 19; DB 15; Length 824;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1491 CAGGGAAGAGATGATGCA 1509
DB      579 CAGGGAAGAGATGATGCA 597
```

```
RESULT 14
US-10-027-632-11812
Sequence 11812, Application US/10027632
```

Job time : 1176.15 secs

```
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108627.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11812
; LENGTH: 824
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11812
```

```
Query Match          1.0%; Score 19; DB 15; Length 824;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1491 CAGGAGAGATGATGCA 1509
          |||||
Db       579 CAGGAGAGATGATGCA 597
```

```
RESULT 15
US-10-369-493-24736
; Sequence 24736, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24736
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
US-10-369-493-24736
```

```
Query Match          1.0%; Score 19; DB 15; Length 1137;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1702 GAGAAAGATGAGAGAGT 1720
          |||||
Db       607 GAGAAAGATGAGAGAGT 625
```

Search completed: March 30, 2004, 23:20:33

**This Page Blank (uspio)**



CC SGR3 results in plants with increased resistance to viruses, while  
CC inactivation of SGR3 by expressing antisense  
CC RNA, by mutation or by homologous recombination, increases the level of  
CC the transgene product. This product may e.g. impart resistance to  
CC herbicide insects or pathogens, alter contents of essential fatty acids  
CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or  
CC interferon

Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 100.0%; Score 1878; DB 4; Length 1878;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGAGTTCTAGGGCTGCTCCATGCTAGAGAAAAGAGCTTCAGAGGTGTTATAGGCT 60
DB 1 ATGAGTTCTAGGGCTGCTCCATGCTAGAGAAAAGAGCTTCAGAGGTGTTATAGGCT 60
QY 61 GAGGTTGAACAGTTGGTTCAAGTTTGGAGGAGCGAGCTGGCTTTTCAAGATGAT 120
DB 61 GAGGTTGAACAGTTGGTTCAAGTTTGGAGGAGCGAGCTGGCTTTTCAAGATGAT 120
QY 121 GAGAGAGAGTGGAGGTCATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 121 GAGAGAGAGTGGAGGTCATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GAGAGAGAGTGGAGGTCATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 121 GAGAGAGAGTGGAGGTCATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 181 AAAAAGTTGGTTCTGAGATTCGAAATCTCTAGAGCTTGGGTTGTCAGAGAGAG 240
DB 181 AAAAAGTTGGTTCTGAGATTCGAAATCTCTAGAGCTTGGGTTGTCAGAGAGAG 240
QY 241 AGAGGTAGCAACGTATCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 241 AGAGGTAGCAACGTATCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 241 AGAGGTAGCAACGTATCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 241 AGAGGTAGCAACGTATCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 301 CGGGGCAATTCAGGTTACATATCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 CGGGGCAATTCAGGTTACATATCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 301 CGGGGCAATTCAGGTTACATATCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 CGGGGCAATTCAGGTTACATATCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAGAGAGAGAGAGAGAGAGAG 420
DB 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAGAGAGAGAGAGAGAGAGAG 420
QY 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAGAGAGAGAGAGAGAGAGAG 420
DB 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAGAGAGAGAGAGAGAGAGAG 420
QY 421 GCAAGAGAGAGTTTCTGCTCAGACACACAGCTGAGAGAGAGAGAGAGAGAGAG 480
DB 421 GCAAGAGAGAGTTTCTGCTCAGACACACAGCTGAGAGAGAGAGAGAGAGAGAG 480
QY 421 GCAAGAGAGAGTTTCTGCTCAGACACACAGCTGAGAGAGAGAGAGAGAGAGAG 480
DB 421 GCAAGAGAGAGTTTCTGCTCAGACACACAGCTGAGAGAGAGAGAGAGAGAGAG 480
QY 481 GTGATTAATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 481 GTGATTAATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 481 GTGATTAATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 481 GTGATTAATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 CTTCGAGAGAGTTTATGATCTCGAGATGAGTCAAAAGAGAGAGAGAGAGAGAGAG 600
DB 541 CTTCGAGAGAGTTTATGATCTCGAGATGAGTCAAAAGAGAGAGAGAGAGAGAGAG 600
QY 541 CTTCGAGAGAGTTTATGATCTCGAGATGAGTCAAAAGAGAGAGAGAGAGAGAGAG 600
DB 541 CTTCGAGAGAGTTTATGATCTCGAGATGAGTCAAAAGAGAGAGAGAGAGAGAGAG 600
QY 601 AATAAGTGGTCAAAAAGTTCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 AATAAGTGGTCAAAAAGTTCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 601 AATAAGTGGTCAAAAAGTTCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 AATAAGTGGTCAAAAAGTTCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 661 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 TATACCTGCAACCTCTAATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TATACCTGCAACCTCTAATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 721 TATACCTGCAACCTCTAATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TATACCTGCAACCTCTAATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 781 CATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 781 CATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 781 CATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 CTTGTGTGAGATTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 CTTGTGTGAGATTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
```

```
QY 901 ATTGCTGGCTCCCAATGATGATCATGATACTAGACTGATTAAGAGAGATACGAT 960
DB 901 ATTGCTGGCTCCCAATGATGATCATGATACTAGACTGATTAAGAGAGATACGAT 960
QY 961 AAGTGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 961 AAGTGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 AGAGCAAGCCTTCTATGATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 AGAGCAAGCCTTCTATGATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1021 AGAGCAAGCCTTCTATGATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 AGAGCAAGCCTTCTATGATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1081 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 TTAGATTAAGATTTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 TTAGATTAAGATTTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1141 TTAGATTAAGATTTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 TTAGATTAAGATTTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 TTAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 TTAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1201 TTAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 TTAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 ACAAGCTGAATTCGAGTTGAATTCATACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 ACAAGCTGAATTCGAGTTGAATTCATACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1261 ACAAGCTGAATTCGAGTTGAATTCATACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 ACAAGCTGAATTCGAGTTGAATTCATACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 ATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 ATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1321 ATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 ATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 CAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 CAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1381 CAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 CAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 GAGGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 GAGGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1441 GAGGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 GAGGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 ATGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1501 ATGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1501 ATGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1501 ATGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1621 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TCAGGCTTCAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1681 TCAGGCTTCAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1681 TCAGGCTTCAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1681 TCAGGCTTCAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 CTGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 CTGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1741 CTGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 CTGATTAATGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 TTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1801 TTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1801 TTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1801 TTTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 CACATTAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1861 CACATTAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1861 CACATTAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1861 CACATTAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
```

RESULT 2  
AAF25373  
ID AAF25373 standard; DNA; 3275 BP.



XX AAF25373;  
AC  
XX  
XX  
DT 15-MAY-2001 (first entry)  
DE Genomic sequence of the Arabidopsis SGS3 gene.  
DE  
XX  
XX  
KW SGS3 gene; post-transcriptional inactivation; RNA degradation;  
KM viral resistance; resistance; fatty acid content; protein content; ss.  
OS  
XX Arabidopsis thaliana.  
XX  
XX  
XX Key Location/Qualifiers  
FH exon 696..1658  
FT /tag= a  
FT 1659..1731  
FT intron /tag= b  
FT 1732..2023  
FT exon /tag= c  
FT /tag= c  
FT intron 2024..2134  
FT /tag= d  
FT exon 2135..2379  
FT /tag= e  
FT intron 2380..2481  
FT /tag= f  
FT exon 2482..2648  
FT /tag= g  
FT intron 2649..2738  
FT /tag= h  
FT exon 2739..2949  
FT /tag= i  
PN WO200105951-A2.  
PN  
PD 25-JAN-2001.  
PD  
XX  
XX 13-JUL-2000; 2000MO-FR002052.  
PF  
XX  
XX 16-JUL-1999; 99FR-00009417.  
PR 26-JAN-2000; 2000FR-00001006.  
PR  
PA (AVET ) AVENTIS CROPS SCIENCE SA.  
PA (INRG ) INST NAT RECH AGRONOMICUE.  
PA  
PI Beclin C, Elmayan T, Vaucheret H;  
PI  
DR WPI; 2001-159529/16.  
DR  
XX  
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
PT resistance in plants and, when inhibited, for increasing transgene  
PT expression.  
PT  
XX  
XX Claim 1; Page 31-32; 36pp; French.  
XX  
XX  
XX The present sequence represents the genomic sequence of the Arabidopsis  
CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional  
CC inactivation (degradation of RNA) and for resistance to viruses.  
CC Overexpression of SGS3 results in plants with increased resistance to  
CC viruses, while inactivation of SGS3 in transgenic plants (e.g. by  
CC expressing antisense RNA, by mutation or by homologous recombination)  
CC increases the level of the transgene product. This product may e.g.  
CC impart resistance (to herbicide, insects or pathogens), alter contents of  
CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an  
CC immunoglobulin or interferon  
XX  
XX Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;

Query Match	51.3%	Score 963	DB 4	Length 3275
Best Local Similarity	100.0%	Pred. No. 0		
Matches 963	Conservative 0	Mismatches 0	Indels 0	Gaps 0

OY      1    ATGAGTTTAGGGCGTTCGCAATGCTCAGAGAAAGAGCTTCAGGGTGTATAGCCCT 60

|||||

Db	696	ATGAGTTCGAGGGCTGGTCCAAATGCTAAAGAAAAGAAAGTTCAAGGTGTATTAGGCT	755
QY	61	GAGGTTGAAACAGTTGGTTCAGAGTTTGGCAGGAGCAGACTGGCTTCTTCAACAATGAT	120
Db	756	GAGGTTGAAACAGTTGGTTCAGAGTTTGGCAGGAGCAGACTGGCTTCTTCAACAATGAT	815
QY	121	GAGGAGAGTGGGAGGTCACTTCCAAAGAAACAAGAACAAACACAGAAACACTTGTGA	180
Db	816	GGAGAGAGTGGGAGGTCACTTCCAAAGAAACAAGAACAAACACAGAAACACTTGTGA	875
QY	181	AAAACTTGGGTTCTCAGAAATTCGAATCCTCCTAGAGCTTGGGGTGTGAGCAGCAAGG	240
Db	876	AAAACTTGGGTTCTCAGAAATTCGAATCCTCCTAGAGCTTGGGGTGTGAGCAGCAAGG	935
QY	241	AGAGGTACCAACGTAATCTGGGAGAGAAACAATGTAATCCGGGAGAGGTAAAGGCATATG	300
Db	936	AGAGGTACCAACGTAATCTGGGAGAGAAACAATGTAATCCGGGAGAGGTAAAGGCATATG	995
QY	301	CGGGGCAATTCAGCTAACATATCTGTCGGGGACGAGCGTTGAGCAGAAAGTATGATAAC	360
Db	996	CGGGGCAATTCAGCTAACATATCTGTCGGGGACGAGCGTTGAGCAGAAAGTATGATAAC	105
QY	361	AACTTTGTGGCACCCCAACCTGTATCTCGCCCTCTTTGGAAGAGATGAAATTTGGCAG	420
Db	1056	AACTTTGTGGCACCCCAACCTGTATCTCGCCCTCTTTGGAAGAGATGAAATTTGGCAG	111
QY	421	GCAAGAGAGAGTTTCTGCTCAGACACACTGTGTCAGAGAGTTTCTGACCTGGAGAGATG	480
Db	1116	GCAAGAGAGAGTTTCTGCTCAGACACACTGTGTCAGAGAGTTTCTGACCTGGAGAGATG	117
QY	481	GTGATAATGCTTCTGAGAGAGAGAAATGTTCCAGATGCTTTGGATGATTTCTATGACGAC	540
Db	1176	GTGATAATGCTTCTGAGAGAGAGAAATGTTCCAGATGCTTTGGATGATTTCTATGACGAC	123
QY	541	CTTGCAGATGATGATTATGACTCGGATGTGAGTCAAAAAGCCATGATCAGGAAAGCAG	600
Db	1236	CTTGCAGATGATGATTATGACTCGGATGTGAGTCAAAAAGCCATGATCAGGAAAGCAG	129
QY	601	AATAAGTGGTTCAAAAAGTTCTTTGGGACCTTGGATCTGTTGATCGATCGACGATTAAT	660
Db	1296	AATAAGTGGTTCAAAAAGTTCTTTGGGACCTTGGATCTGTTGATCGACGATTAAT	135
QY	661	GAACCAACAGAGGACAGTGGCACTTGTCCAGAACGCACTTGATGCCATTCGATTGG	720
Db	1356	GAACCAACAGAGGACAGTGGCACTTGTCCAGAACGCACTTGATGCCATTCGATTGG	141
QY	721	TATAAAGCTGACCCCTTACTAGCTCATGCGAGCAAAAAGAGCTAGGCGAGTTAAGCTC	780
Db	1416	TATAAAGCTGACCCCTTACTAGCTCATGCGAGCAAAAAGAGCTAGGCGAGTTAAGCTC	147
QY	781	CATAGAGAAATGGCTGAGATTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGTCAAT	840
Db	1476	CATAGAGAAATGGCTGAGATTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGTCAAT	153
QY	841	CCTTGATGATGATTTATGAGCAGTGAAGGGTTGGGTGAGAGATGAAGAAAGATTAAGAA	900
Db	1536	CCTTGATGATGATTTATGAGCAGTGAAGGGTTGGGTGAGAGATGAAGAAAGATTAAGAA	159
QY	901	ATTGTCTGGCCTCCAAATGTCATCATCTGAAATCTAGACTGAGATTAAGACGATACGAT	960
Db	1596	ATTGTCTGGCCTCCAAATGTCATCATCTGAAATCTAGACTGAGATTAAGACGATACGAT	165
QY	961	AAG 963	
Db	1656	AAG 1658	

RESULT	3
AAAF25372/c	
ID	AAAF25372 standard; DNA; 27 BP
XX	
AC	AAAF25372;
XX	

DT 15-MAY-2001 (first entry)  
 XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.  
 DE  
 XX  
 XX SGS3 gene; post-transcriptional inactivation; RNA degradation;  
 KM viral resistance; resistance; fatty acid content; protein content;  
 KM PCR primer; ss.  
 XX  
 XX Arabidopsis thaliana.  
 OS  
 XX WO200105951-A2.  
 PN  
 XX 25-JAN-2001.  
 PD  
 XX 13-JUL-2000; 2000WO-FR002052.  
 PF  
 XX 16-JUL-1999; 99FR-00009417.  
 PR 26-JAN-2000; 2000FR-00001006.  
 XX  
 XX (AVERT ) AVENTIS CROPS SCIENCE SA.  
 PA (INRG ) INST NAT RECH AGRONOMIQUE.  
 XX  
 XX Beclin C, Elmayer T, Vaucheret H;  
 PI WPI; 2001-159529/16.  
 DR  
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 PT resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 PT  
 XX  
 XX Example 1; Page 22; 36pp; French.  
 PS  
 XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis  
 CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-  
 CC transcriptional inactivation (degradation of RNA) and for resistance  
 CC to viruses. Overexpression of SGS3 results in plants with increased  
 CC resistance to viruses, while inactivation of SGS3 in transgenic plants  
 CC (e.g. by expressing antisense RNA, by mutation or by homologous  
 CC recombination) increases the level of the transgene product. This product  
 CC may e.g. impart resistance (to herbicide, insects or pathogens), alter  
 CC contents of essential fatty acids or proteins, or is pharmaceutically  
 CC active, e.g. an immunoglobulin or interferon  
 CC  
 XX Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;  
 SQ  
 Query Match 1.3%; Score 24; DB 4; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 0.71;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1855 GGCCTTCACATGAGATGATGA 1878  
 DB 27 GGCCTTCACATGAGATGATGA 4  
 RESULT 4  
 AAF25371  
 ID AAF25371 standard; DNA; 23 BP.  
 XX  
 XX AAF25371;  
 AC  
 XX  
 XX 15-MAY-2001 (first entry)  
 DT  
 XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.  
 DE  
 XX  
 XX SGS3 gene; post-transcriptional inactivation; RNA degradation;  
 KM viral resistance; resistance; fatty acid content; protein content;  
 KM PCR primer; ss.  
 XX  
 XX Arabidopsis thaliana.  
 OS  
 XX WO200105951-A2.  
 PN  
 XX 25-JAN-2001.  
 PD

XX  
 PF 13-JUL-2000; 2000WO-FR002052.  
 XX  
 XX 16-JUL-1999; 99FR-00009417.  
 PR 26-JAN-2000; 2000FR-00001006.  
 XX  
 XX (AVERT ) AVENTIS CROPS SCIENCE SA.  
 PA (INRG ) INST NAT RECH AGRONOMIQUE.  
 XX  
 XX Beclin C, Elmayer T, Vaucheret H;  
 PI WPI; 2001-159529/16.  
 DR  
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 PT resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 PT  
 XX  
 XX Example 1; Page 22; 36pp; French.  
 PS  
 XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis  
 CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-  
 CC transcriptional inactivation (degradation of RNA) and for resistance  
 CC to viruses. Overexpression of SGS3 results in plants with increased  
 CC resistance to viruses, while inactivation of SGS3 in transgenic plants  
 CC (e.g. by expressing antisense RNA, by mutation or by homologous  
 CC recombination) increases the level of the transgene product. This product  
 CC may e.g. impart resistance (to herbicide, insects or pathogens), alter  
 CC contents of essential fatty acids or proteins, or is pharmaceutically  
 CC active, e.g. an immunoglobulin or interferon  
 CC  
 XX Sequence 23 BP; 6 A; 4 C; 7 G; 6 T; 0 U; 0 Other;  
 SQ  
 Query Match 1.1%; Score 20; DB 4; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ATGACTTTAGCGCTGCTCC 20  
 DB 4 ATGACTTTAGCGCTGCTCC 23  
 RESULT 5  
 AAF12773  
 ID AAF12773 standard; cDNA; 596 BP.  
 XX  
 XX AAF12773;  
 AC  
 XX  
 XX 13-MAR-2001 (first entry)  
 DT  
 XX Aspergillus oryzae EST SEQ ID NO:5296.  
 DE  
 XX  
 XX Multiple gene expression; filamentous fungal cell; EST;  
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KM culture condition; environmental stress; spore morphogenesis;  
 KM metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 XX Aspergillus oryzae.  
 OS  
 XX WO200056762-A2.  
 PN  
 XX 28-SEP-2000.  
 PD  
 XX 22-MAR-2000; 2000WO-US007781.  
 PF  
 XX 22-MAR-1999; 99US-00273623.  
 PR  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 XX Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB,  
 PI WPI; 2000-594572/56.  
 DR

XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags.  
PS Claim 88; Page 2210; 3161pp; English.  
XX  
XX The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring the  
CC global expression of genes from FF cells allows the production potential  
CC of the microorganisms to be improved. New genes may be discovered,  
CC possible functions of unknown open reading frames can be identified and  
CC gene copy number variation and stability can be monitored. The expression  
CC of genes can be used to study how FF cells adapt to changes in culture  
CC conditions, environmental stress, spore morphogenesis, recombination,  
CC metabolic or catabolic pathway engineering. Using ESTs provides several  
CC advantages over genomic or random cDNA clones including elimination of  
CC redundancy as one spot on an array equals one gene or open reading frame,  
CC and organization of the microarrays based on function of the gene  
CC products to facilitate analysis of the results. AA07478 to AA01247  
CC represents ESTs from *Fusarium venenatum*; AA01248 to AA01853 represents  
CC ESTs from *Aspergillus niger*; AA01854 to AA014878 represents ESTs from  
CC *Aspergillus oryzae*; and AA014879 to AA01537 represents ESTs from  
CC *Trichoderma reesei*, which are all specifically claimed in the present  
CC invention  
XX  
SQ Sequence 596 BP; 144 A; 168 C; 150 G; 134 T; 0 U; 0 Other;  
Query Match 1.1%; Score 20; DB 3; Length 596;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 142 TCCAGAGACAGACAGACAA 161  
|||  
431 TCCAGAGACAGACAGACAA 450  
DB

RESULT 6  
AA093017/c  
ID AA093017 standard; DNA; 700 BP.  
XX  
XX AA093017;  
AC  
XX  
XX 09-OCT-2001 (first entry)  
DT  
XX  
XX Human inflammatory bowel disease related gene fragment IGR3310a.  
DE  
XX  
XX Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
XX chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200142511-A2.  
XX  
XX 14-JUN-2001.  
XX  
XX 11-DEC-2000; 2000WO-US033632.  
PF  
XX  
XX 10-DEC-1999; 99US-0170257P.  
PR  
XX 10-APR-2000; 2000US-0196046P.  
XX  
XX (WHD) WHITEHEAD INST BIOMEDICAL RES.  
PA (ELI-) ELIIPSIS BIOTHERAPEUTICS CORP.  
XX  
XX Daily M, Hudson TV, Lander ES, Rioux J, Siminovitch K;  
PI

DR WPI; 2001-367874/38.  
XX  
XX Testing for the presence of polymorphisms associated with inflammatory  
PT bowel disease, using a hybridization assay.  
XX  
XX Disclosure; Page 381; 463pp; English.  
XX  
XX The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
CC invention  
XX  
SQ Sequence 700 BP; 156 A; 158 C; 149 G; 237 T; 0 U; 0 Other;  
Query Match 1.1%; Score 20; DB 4; Length 700;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1696 TTCAAGAGAAAGAGATGCA 1715  
|||  
415 TTCAAGAGAAAGAGATGCA 396  
DB

RESULT 7  
AA081465/c  
ID AA081465 standard; DNA; 47475 BP.  
XX  
XX AA081465;  
AC  
XX  
XX 04-DEC-2000 (first entry)  
DT  
XX  
XX N. meningitidis partial DNA sequence gnm\_13 SEQ ID NO:13.  
DE  
XX  
XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
XX *Meningococcus B*; MenB; ds.  
XX  
XX *Neisseria meningitidis*.  
XX  
XX WO200022430-A2.  
XX  
XX 20-APR-2000.  
PD  
XX  
XX 08-OCT-1999; 99WO-US023573.  
PF  
XX  
XX 09-OCT-1998; 98US-0103794P.  
PR  
XX 30-APR-1999; 99US-0132068P.  
XX  
XX (CHTR) CHIRON CORP.  
PA  
XX  
XX Frazer CM, Hickey E, Peterson J, Tetzelin H, Venter JC;  
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarelli M, Scarlato V;  
PI Rapinotti R, Pizsa M;  
XX  
XX WPI; 2000-318079/27.  
XX  
XX  
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used  
PT in the diagnosis and treatment of *N. meningitidis* infection and other  
PT *Neisseria* infections, for example, *N. gonorrhoea*.  
XX  
XX Claim 7; Page 383-397; 1760pp; English.  
XX  
XX The present invention describes methods of obtaining immunogenic proteins  
CC from *Neisseria* genomic sequences. AA081453 to AA082414 represent  
CC specifically claimed *Neisseria meningitidis* genomic DNA sequences;  
CC AA081260 to AA081303 and AA0825620 to AA0825663 represent *Neisseria* DNA  
CC sequences and their corresponding proteins; AA081254 to AA081259 and  
CC AA081304 to AA081321 represent PCR primers used in the isolation of

CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent  
 CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all  
 CC used in the exemplification of the present invention. The nucleic acid  
 CC sequences, protein sequences, and antibodies against them, can be used in  
 CC the manufacture of a composition. The composition can be used as a  
 CC medicament (or in the manufacture of a medicament) for treating,  
 CC preventing or diagnosing infection due to Neisserial bacteria. For  
 CC example, some of the identified proteins could be components of vaccines  
 CC against Meningococcus B; against all serotypes; and/or against all  
 CC pathogenic Neisseriae. Identification of biological probes, particularly  
 CC will also facilitate production of biological probes, particularly  
 CC organism-specific probes. Attempts to make efficacious Meningococcus B  
 CC vaccines have also been tried but none have successfully overcome  
 CC antigenic variability. The provision of further, complete sequences may  
 CC provide an opportunity to identify secreted or surface exposed proteins  
 CC that may be presumed targets for the immune system and which are not  
 CC antigenically variable or at least more conserved than other more  
 CC variable regions

XX  
 XX Sequence 47475 BP; 11600 A; 13758 C; 12290 G; 9826 T; 0 U; 1 Other;

XX  
 XX Query Match 1.1%; Score 20; DB 3; Length 47475;  
 XX Best Local Similarity 100.0%; Pred. No. 55;  
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX  
 XX 67 GAACAGTTGGTTCAGGTTT 86  
 XX 23354 GAACAGTTGGTTCAGGTTT 23335

XX  
 XX RESULT 8  
 XX AAA81489 5/c  
 XX Continuation (6 of 9) of AAA81489 from base 500001 (N. meningitidis partial DNA sequence  
 XX WP Sequence split into 9 fragments LOCUS AAA81489 Accession Aaa81489  
 XX Fragment Name Begin End  
 XX WP AAA81489\_0 1 110000  
 XX WP AAA81489\_1 100001 210000  
 XX WP AAA81489\_2 200001 310000  
 XX WP AAA81489\_3 300001 410000  
 XX WP AAA81489\_4 400001 510000  
 XX WP AAA81489\_5 500001 610000  
 XX WP AAA81489\_6 600001 710000  
 XX WP AAA81489\_7 700001 810000  
 XX WP AAA81489\_8 800001 837096

XX  
 XX Query Match 1.1%; Score 20; DB 3; Length 110000;  
 XX Best Local Similarity 100.0%; Pred. No. 55;  
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX  
 XX 67 GAACAGTTGGTTCAGGTTT 86  
 XX 32595 GAACAGTTGGTTCAGGTTT 32576

XX  
 XX RESULT 9  
 XX ABO88198/c  
 XX ID ABO88198 standard; cDNA; 154902 BP.  
 XX AC ABO88198;  
 XX DT 18-SEP-2002 (first entry)  
 XX DE Human osteoblast differentiation related cDNA SEQ ID NO 105.  
 XX DE Human; osteoblast; stem cell differentiation; bone tissue deposition;  
 XX DE osteoporosis; osteopathic; ss.  
 XX OS Homo sapiens.  
 XX PN WO200250301-A2.  
 XX PD 27-JUN-2002.

XX  
 XX 18-DEC-2001; 2001MO-US048276.  
 XX 18-DEC-2000; 2000US-0255882P.  
 XX 24-APR-2001; 2001US-0285691P.  
 XX (GENE-) GENE LOGIC INC.  
 XX (PROC) PROCTER & GAMBLE CO.  
 XX JI D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
 XX Mertz L;  
 XX WPI; 2002-557663/59.  
 XX Use of genes and their expression profiles associated with osteoblast  
 XX differentiation for screening modulators bone formation, for diagnosing  
 XX or treating e.g. osteoporosis, or as markers for the differentiation  
 XX process.

XX  
 XX Claim 1; SEQ ID NO 105; 78pp + Sequence Listing; English.

XX  
 XX The invention relates to genes and their expression profiles are used  
 XX for: (a) screening modulators of precursor stem cell differentiation into  
 XX osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
 XX deposition of bone tissue, abnormal rate of osteoblast formation or  
 XX osteoporosis; or (c) treating or monitoring treatment of the conditions  
 XX cited in (b) or monitoring the progression of bone tissue deposition.  
 XX Specific conditions include postmenopausal osteoporosis, glucocorticoid  
 XX induced osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-  
 XX induced abnormalities in bone formation or bone loss, conditions that  
 XX involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
 XX skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
 XX or fibrous dysplasia. The present sequence is that of an osteoblast  
 XX differentiation associated cDNA marker of the invention. Note: The  
 XX sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences

XX  
 XX Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;

XX  
 XX Query Match 1.1%; Score 20; DB 6; Length 154902;  
 XX Best Local Similarity 100.0%; Pred. No. 54;  
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX  
 XX 1712 TGGAGAGTTTGTGGAAG 1731  
 XX 128153 TGGAGAGTTTGTGGAAG 128134

XX  
 XX RESULT 10  
 XX AAF21612/c  
 XX ID AAF21612 standard; DNA; 349980 BP.  
 XX AC AAF21612;  
 XX DT 13-MAR-2001 (first entry)  
 XX DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.  
 XX DE Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 XX DE diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 XX ds.  
 XX OS Neisseria meningitidis.  
 XX PN WO2000066791-A1.  
 XX PD 09-NOV-2000.  
 XX 08-MAR-2000; 2000MO-US005928.  
 XX 30-APR-1999; 99US-0132068P.  
 XX 08-OCT-1999; 99MO-US023573.

PR 28-FEB-2000; 2000GB-00004695.  
XX (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Pizza M, Hickey E, Peterson J, Tetelín H, Venter JC,  
PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V,  
PI Rappuoli R, Frazer CM, Grandi G;  
XX WPI; 2000-647603/62.  
DR  
XX  
XX Neisseria meningitidis B full length genome sequence and open reading  
PT frames are used to detect, treat and prevent Neisserial infections.  
XX  
XX Claim 7; Appendix A; 692pp; English.  
XX  
XX The present invention describes the full length genome of Neisseria  
CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613  
CC represent fragments of the NMB genomic sequence, as the sequence was too  
CC long to go in a record on its own it was split into 8 sequences which  
CC overlap each other at the beginning and end of each sequence by 49980 bp  
CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of  
CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of  
CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins  
CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR  
CC primers which are used in the exemplification of the present invention.  
CC The NMB genome and fragments from it have antibacterial activity, and can  
CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins  
CC and/or antibodies which binds to the proteins can be used in compositions  
CC for treating or preventing infection due to Neisserial bacteria or as a  
CC diagnostic reagent for detecting the presence of Neisserial bacteria or  
CC of antibodies raised to Neisserial bacteria. Computers, computer memory,  
CC computer storage medium or computer databases can be used in a search to  
CC identify open reading frames (ORFs) or coding sequences within the NMB  
CC genome. The DNA sequences provide further opportunities to find antigenic  
CC or immunogenic proteins which are more effective in vaccines than the  
CC outer membrane proteins currently used  
XX  
SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;  
XX  
Query Match 1.1%; Score 20; DB 3; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 67 GAACAGTGGTCAAGTTT 86  
Db 167824 GAACAGTGGTCAAGTTT 167805  
XX  
RESULT 11  
AAA68016/C  
ID AAA68016 standard; DNA; 481 BP.  
XX  
XX AAA68016;  
AC  
XX  
DT 24-OCT-2000 (first entry)  
XX  
XX Eucalyptus grandis CCR nucleotide sequence SEQ ID NO:109.  
DE  
XX  
XX Eucalyptus grandis CCR nucleotide sequence SEQ ID NO:109.  
DE  
XX  
XX Plant; lignin; lignin biosynthetic pathway; Eucalyptus grandis;  
KW Pinus radiata; Monterey pine; ds.  
XX  
XX Eucalyptus grandis.  
OS  
XX  
XX WO200022099-A1.  
FN  
XX  
XX 20-APR-2000.  
PD  
XX  
XX 06-OCT-1999; 99WO-NZ000168.  
PF  
XX  
XX 09-OCT-1998; 98US-00169789.  
PR  
XX 14-JUL-1999; 99US-0143811P.  
PR  
XX

PA (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
XX  
PI Bloksberg LN, Havukkala ID;  
XX  
DR WPI; 2000-317962/27.  
XX  
XX Novel polynucleotide encoding enzymes involved in lignin-biosynthetic  
PT pathway useful for producing transgenic plants especially eucalyptus and  
PT pine species having altered lignin content, composition and structure.  
XX  
XX  
PS Claim 1; Page 91-92; 213pp; English.  
XX  
XX  
XX The present invention describes isolated polynucleotides and proteins  
CC encoding and representing the enzymes cinnamoyl 4-hydroxylase (CH),  
CC coumarate 3-hydroxylase (CH), phenolase (PNU), O-methyl transferase  
CC (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase  
CC (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL),  
CC coniferol glucosyl transferase (CGT), coniferin beta-glucosidase (CBG),  
CC laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase, caffeic  
CC acid methyl transferase, caffeoyl CoA methyl transferase, coumarate CoA  
CC ligase, cytochrome P450 1X1A, diphenol oxidase, flavanol glucosyl  
CC transferase, flavonoid hydroxylase, and isoflavone reductase, which are  
CC involved in the lignin biosynthetic pathway. The polynucleotides can be  
CC used for modulating lignin content, lignin composition and the structure  
CC of a plant, especially eucalyptus and pine species, and for modifying the  
CC activity of an enzyme involved in lignin biosynthetic pathway, and for  
CC producing a plant having altered lignin content, composition and  
CC structure. They can be used for designing probes and primers useful for  
CC detecting similar DNA and RNA sequences in any organism and for PCR  
CC amplification. The lignin content can be efficiently modified using the  
CC polynucleotides. AAB6798 to AAB68201 and AAB16341 to AAB16449 represent  
CC polynucleotide and protein sequences used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 481 BP; 122 A; 118 C; 111 G; 130 T; 0 U; 0 Other;  
XX  
Query Match 1.0%; Score 19; DB 3; Length 481;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 TGGTCAAGTTTGGCAGG 92  
Db 470 TGGTCAAGTTTGGCAGG 452  
XX  
RESULT 12  
ADD41766/C  
ID ADD41766 standard; DNA; 481 BP.  
XX  
XX ADD41766;  
AC  
XX  
DT 15-JAN-2004 (first entry)  
XX  
XX Cinnamoyl-CoA reductase DNA #21.  
DE  
XX  
XX ds; plant; lignin content; lignin composition; lignin structure;  
KW lignin biosynthesis; gene.  
XX  
XX Eucalyptus grandis.  
OS  
XX  
XX US2003131373-A1.  
FN  
XX  
XX 10-JUL-2003.  
PD  
XX  
XX 18-JUN-2002; 2002US-00174693.  
PF  
XX  
XX 11-SEP-1996; 96US-00713000.  
PR  
XX 21-NOV-1997; 97US-00975316.  
PR  
XX 09-OCT-1998; 98US-00169789.  
PR  
XX 14-JUL-1999; 99US-0143833P.  
PR  
XX 12-JUL-2000; 2000US-00615192.  
XX

PA (BLOC/) BLOKSBERG L N.  
PA (HAYU/) HAYUKKALA I.  
XX BLOKSBERG LN, HAYUKKALA I;  
XX WPI, 2003-829606/77.  
XX New polynucleotide associated with the lignin biosynthetic pathway, for  
XX modulating lignin content, composition and structure of plants, or  
XX PT producing a plant with altered lignin content, composition and structure.  
XX  
XX Claim 1, SEQ ID NO 109, 206pp; English.  
XX  
XX The invention relates to an isolated polynucleotide. The polynucleotides,  
XX CC polypeptides and genetic constructs are useful for modulating lignin  
XX CC content, composition and structure of plants, or for producing a plant  
XX CC having altered lignin content, composition and structure. The  
XX CC polynucleotides are also useful in genome mapping, physical mapping,  
XX CC positional cloning of genes, or as non-disruptive tags for marking  
XX CC organisms, particularly plants. The present sequence represents a  
XX CC polynucleotide associated with the lignin biosynthetic pathway.  
XX  
XX Sequence 481 BP, 122 A, 118 C, 111 G, 130 T, 0 U, 0 Other;  
SQ  
Query Match 1.0%; Score 19; DB 9; Length 481;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 TGGTTCAAGGTTGGCAGG 92  
Db 470 TGGTTCAAGGTTGGCAGG 452  
RESULT 13  
AAK64246  
ID AAK64246 standard; CDNA; 484 BP.  
XX AAK64246;  
XX AC  
XX AAK64246;  
XX DT 06-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:9306.  
DE Human immune/haematopoietic; immune/haematopoietic antigen; cancer;  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157182-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 17-JAN-2001, 2001WO-US001354.  
PF  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-018650P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 01-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 05-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 06-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 08-SEP-2000; 2000US-0231968P.  
PR 12-SEP-2000; 2000US-0231969P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 14-SEP-2000; 2000US-0233066P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 26-SEP-2000; 2000US-0235485P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 29-SEP-2000; 2000US-0236371P.  
PR 29-SEP-2000; 2000US-0236372P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 02-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 13-OCT-2000; 2000US-0239338P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249255P.  
 PR 17-NOV-2000; 2000US-0249257P.  
 PR 17-NOV-2000; 2000US-0249259P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251888P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251899P.  
 PR 08-DEC-2000; 2000US-0251909P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI: 2001-483426/52.  
 DR P-PSDB; AAM91465.  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
 PS Claim 1, SEQ ID NO 9306; 3071pp + Sequence Listing; English.

CC represent sequences used in the exemplification of the present invention  
 XX Sequence 484 BP; 179 A; 66 C; 106 G; 131 T; 0 U; 2 Other;  
 SQ  
 Query Match 1.0%; Score 19; DB 4; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1767 GGAGACATGAGAGAGG 1785  
 |||||  
 DB 323 GGAGACATGAGAGAGG 341  
 RESULT 14  
 AAV23847/c  
 ID AAV23847 standard; DNA; 495 BP.  
 XX AAV23847;  
 AC  
 XX  
 DT 31-JUL-1998 (first entry)  
 XX  
 DE Plant CCR enzyme DNA sequence.  
 XX  
 KW Lignin biosynthetic pathway; eucalyptus; pine; transgenic plant;  
 KW lignin content; tree processing; cellulose fibre; ss.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 PN WO9811205-A2.  
 XX  
 PD 19-MAR-1998.  
 XX  
 PF 10-SEP-1997; 97WO-NZ000112.  
 XX  
 PR 11-SEP-1996; 96US-00713000.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 PI Bloksberg LN, Grierson AW, Havukkala IU;  
 DR WPI: 1998-207374/18.  
 XX  
 PT Sequences useful for modification of plant lignin content or structure -  
 PT from Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) are  
 PT associated with lignin biosynthesis pathway, useful e.g. in paper  
 PT industry.  
 PS Claim 1; Page 39; 82pp; English.

CC This sequence represents a fragment of the CCR enzyme coding sequence. It  
 CC is an example of a DNA sequence of the invention, which are from  
 CC Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) associated with  
 CC the lignin biosynthesis pathway. Constructs containing the DNA sequences  
 CC can be used to produce transgenic plants or plant cells, especially woody  
 CC plants e.g. eucalyptus or pine species but also e.g. monocotyledons or  
 CC dicotyledons; by stably incorporating the constructs into the plant  
 CC genome. The lignin content or structure, or activity of a specific enzyme  
 CC in the plant, can therefore be modulated. Reductions in lignin content or  
 CC changes in composition are useful in tree processing for paper. High  
 CC lignin content results in energy- and chemical-intensive separation  
 CC methods in order to obtain the pure cellulose fibre required. Reductions  
 CC in lignin content may also be useful for forage crops, whilst increases  
 CC or changes in composition may be desirable to increase the mechanical  
 CC strength of wood, change its colour or increase its resistance to rot.  
 CC The sequences are also useful as probes to isolate DNA sequences encoding  
 CC enzymes involved in the lignin biosynthesis pathway from other plant  
 CC species  
 XX  
 SQ Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;  
 Query Match 1.0%; Score 19; DB 2; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;



Search completed: March 30, 2004, 09:26:50  
Job time : 655.082 secs

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 74 TGGTTCAAGTTTGGCAG 92  
 |||||  
 Db 484 TGGTTCAAGTTTGGCAG 466

RESULT 15  
 AA206848/C  
 ID AA206848 standard; cDNA; 495 BP.

AC AA206848;

DT 09-NOV-1999 (first entry)

DE Eucalyptus cinamoyl-CoA reductase (CCR) partial cDNA 2.

KM lignin; biosynthesis; forage crop; wood; paper production;

KW transgenic plant; 58.

XX Eucalyptus grandis.

OS

PN US5952486-A.

PD 14-SEP-1999.

PF 21-NOV-1997; 97US-00975316.

PR 11-SEP-1996; 96US-00713000.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Grierson AW, Bloksberg LN, Havukkala I;

XX WPI; 1999-527029/44.

DR Isolated DNA sequence encoding enzymes from the lignin synthetic pathway

XX useful for generating plants with an altered lignin content.

PS Example 1; Col 37-38; 48pp; English.

XX This sequence represents a cinamoyl-CoA reductase (CCR) partial cDNA

CC from Eucalyptus grandis. This enzyme is involved in the biosynthesis of

CC lignin, an insoluble polymer which is primarily responsible for the

CC rigidity of plant stems. Lignin serves as a matrix around the

CC polysaccharide components of some plant cell walls. The higher the lignin

CC content, the more rigid the plant. Lignin also plays a role in disease

CC resistance of plants by impeding the penetration and propagation of

CC pathogenic agents. Lignin is formed by polymerisation of at least three

CC different monolignols (para-coumaryl alcohol, coniferyl alcohol and

CC sinapyl alcohol). These three monolignols are synthesised by similar

CC pathways from phenylalanine in a multistep process and are believed to be

CC polymerised into lignin via a free radical mechanism. The lignin content

CC of plants can be altered using DNA sequences encoding these enzymes.

CC Lignin content can be increased by the target plant. This could be

CC beneficial for increasing the mechanical strength of wood. Similarly, a

CC decrease in lignin content can be obtained by transforming the target

CC plant with antisense copies of such genes. This may be beneficial in

CC plant with antisense copies of such genes. This may be beneficial in

CC trees used as forage crops for livestock (lignin is indigestible) and in

CC trees used in paper manufacture

XX

XX Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;

XX

XX Query Match 1.0%; Score 19; DB 2; Length 495;

XX Best Local Similarity 100.0%; Pred. No. 1.8e+02;

XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 74 TGGTTCAAGTTTGGCAG 92

XX |||||

XX Db 484 TGGTTCAAGTTTGGCAG 466

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 03:15:12 ; Search time 7026.56 Seconds

(without alignments)  
11584.370 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878  
Sequence: 1 atgagctctagagctgctgctc.....cttcacatgaagatgatcga 1878Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516395 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

```
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_ey:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1878	100.0	1878	6	AX078761	AX078761 Sequence
2	1878	100.0	1909	8	BT004380	BT004380 Arabidops
3	1878	100.0	2162	8	BT002944	BT002944 Arabidops
4	963	51.3	2254	8	AF239719	AF239719 Arabidops
5	963	51.3	3275	6	AX078760	AX078760 Sequence
6	963	51.3	81365	8	AB025633	AB025633 Arabidops
7	239	12.7	650	8	ATH528171	ATH528171 Arabidops
8	22	1.2	179828	2	AC121725	AC121725 Rattus no
9	22	1.2	179709	10	AL928719	AL928719 Mouse DNA
10	22	1.2	214186	2	AC078933	AC078933 Mus muscu
11	22	1.2	218774	2	AC079182	AC079182 Mus muscu
12	22	1.2	236508	2	AC115306	AC115306 Rattus no
13	22	1.2	280016	2	AC140364	AC140364 Mus muscu
14	22	1.2	304407	2	AC073813	AC073813 Mus muscu
15	21	1.1	10225	1	AE005048	AE005048 Halobacte
16	21	1.1	50685	9	AL139004	AL139004 Human DNA
17	21	1.1	76545	2	AC080117	AC080117 Homo sapi
18	21	1.1	79554	8	NCB11823	NCB11823 Neurospor
19	21	1.1	110603	9	HS421D16	HS421D16 Human DNA
20	21	1.1	110654	2	AC146790	AC146790 Medicago
21	21	1.1	124687	9	AL592164	AL592164 Human DNA
22	21	1.1	132029	9	HS795G23	HS795G23 Homo sapi
23	21	1.1	136932	9	AC068035	AC068035 Homo sapi
24	21	1.1	143051	1	D90911	D90911 Synchocyst
25	21	1.1	160658	9	AC024258	AC024258 Homo sapi
26	21	1.1	173751	2	AL590292	AL590292 Homo sapi
27	21	1.1	175265	2	AC063964	AC063964 Homo sapi
28	21	1.1	180707	2	AC024460	AC024460 Homo sapi
29	21	1.1	183283	2	AC131075	AC131075 Mus muscu
30	21	1.1	200254	2	AC135949	AC135949 Homo sapi
31	21	1.1	202589	2	AC102629	AC102629 Mus muscu
32	21	1.1	211661	2	AC117324	AC117324 Rattus no
33	21	1.1	215773	9	AC126121	AC126121 Homo sapi
34	21	1.1	230654	2	AC115968	AC115968 Mus muscu
35	21	1.1	232071	2	AC120347	AC120347 Mus muscu
36	21	1.1	234516	2	AC108246	AC108246 Rattus no
37	21	1.1	241298	2	AC127667	AC127667 Rattus no
38	21	1.1	271898	2	AC106311	AC106311 Rattus no
39	21	1.1	300242	1	AE016790	AE016790 Pseudomon
40	20	1.1	428	8	AY135509	AY135509 Cestrum a
41	20	1.1	700	6	AX183276	AX183276 Sequence
42	20	1.1	2530	8	AK065651	AK065651 Oryza sat
43	20	1.1	4310	10	AF134918	AF134918 Mus muscu
44	20	1.1	4326	10	BC049183	BC049183 Mus muscu
45	20	1.1	4460	10	BC054532	BC054532 Mus muscu

## ALIGNMENTS

```
RESULT 1
LOCUS AX078761 1878 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 2 from Patent WO01005951.
ACCESSION AX078761
VERSION AX078761.1 GI:13158380
SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS Beclin,C., Elmayan,T. and Vaucheret,H.
TITLE Novel sgsl plant gene and use thereof
```



Db 1681 TCAGCTTCATCGAGTTTCCAGAGAAAGATGAGAGAGTTTGTGAGAGAGGAGAGG 1740  
Qy 1741 CTGATTAAGATCAAGAGAAAGATGAGAGATGAGAGATGAGAGAGGATGAGAGATGATA 1800  
Db 1741 CTGATTAAGATCAAGAGAAAGATGAGAGATGAGAGATGAGAGAGGATGAGAGAGATA 1800  
Qy 1801 TTTGATCTGAGAGAAAGATTTTGTATGAGGCTTTGGAACAGCTCATGTACAGATGACCTT 1860  
Db 1801 TTTGATCTGAGAGAAAGATTTTGTATGAGGCTTTGGAACAGCTCATGTACAGATGACCTT 1860  
Qy 1861 CACATGAAGATGATGA 1878  
Db 1861 CACATGAAGATGATGA 1878  
RESULT 2  
BT004380 1909 bp mRNA linear PLN 14-FEB-2003  
LOCUS Arabidopsis thaliana clone U20243 unknown protein (At5g23570) mRNA,  
DEFINITION complete cds.  
ACCESSION BT004380.1 GI:28393932  
KEYWORDS FL1 cDNA.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 1909)  
REFERENCE  
AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,  
Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W.,  
Ecker,J.R. and Theologis,A.  
Arabidopsis Open Reading Frame (ORF) Clones  
Unpublished  
2 (bases 1 to 1909)  
Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,  
Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W.,  
Ecker,J.R. and Theologis,A.  
Direct Submission  
Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
The RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RFL CDNA (RFL CDNA : RIKEN  
Arabidopsis Full-Length CDNA): Seki,M., Narusaka,M., Ishida,J.,  
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
Hayashizaki,Y. and Shinzaki,K.  
The Sak, Stanford, PGEC (SSP) Consortium members constructed and  
sequenced the PUNI (ORF) clones using the RFL CDNA: Yamada,K.,  
Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,  
Palm,C.J., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,  
Ecker,J.R. and Theologis,A.  
Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally  
to this work. Shinzaki,K. (RIKEN GSC) and Theologis,A. (SSP  
/PGEC) contributed equally to this work as PIs.  
Annotation based on July 2002 version of the Arabidopsis genome  
submitted to Genbank.  
FEATURES  
source  
1..1909  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"

3'UTR  
ORIGIN  
Query Match 100.0%; Score 1878; DB 8; Length 1909;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ATGAGTTCTAGGCTGTGCTCAATGCTTAAAGAAAGACCTTCAGGCTTATAGGCTT 60  
1 ATGAGTTCTAGGCTGTGCTCAATGCTTAAAGAAAGACCTTCAGGCTTATAGGCTT 60  
Db 1 ATGAGTTCTAGGCTGTGCTCAATGCTTAAAGAAAGACCTTCAGGCTTATAGGCTT 60  
Qy 61 GAGGTTGAACATGCTTCAAGTTTGGCAGGAGAGAGCTGCTTCAAGATGAT 120  
Db 61 GAGGTTGAACATGCTTCAAGTTTGGCAGGAGAGAGCTGCTTCAAGATGAT 120  
Qy 121 GGAGGAGAGTGGAGAGTCTTCCAAAGAAAGACCAAGAACCAAGAACCTTCTGGA 180  
Db 121 GGAGGAGAGTGGAGAGTCTTCCAAAGAAAGACCAAGAACCAAGAACCTTCTGGA 180  
Qy 181 AAACTGTGCTTCTCAGATTCGATCTCTCTAGAGCTTGGGCTGTGTCAGAGCAAGG 240  
Db 181 AAACTGTGCTTCTCAGATTCGATCTCTCTAGAGCTTGGGCTGTGTCAGAGCAAGG 240  
Qy 241 AGAGGTAGCAAGCTATCTGGAGAGAGAAACATATATCTGGAGAGAGTAAACGCAATG 300  
Db 241 AGAGGTAGCAAGCTATCTGGAGAGAGAAACATATATCTGGAGAGAGTAAACGCAATG 300  
Qy 301 CGGGGCAATCAAGCTAATCATCTGCGGGAGAGAGGCTTGGAGCAAGATATGATAAC 360  
Db 301 CGGGGCAATCAAGCTAATCATCTGCGGGAGAGAGGCTTGGAGCAAGATATGATAAC 360  
Qy 361 AACTTTGTGACACCCCACTGATCTGCGCTCTTTGGAAGAGAGATGATGAGCAG 420  
Db 361 AACTTTGTGACACCCCACTGATCTGCGCTCTTTGGAAGAGAGATGATGAGCAG 420  
Qy 421 GCAAGAGAGGTTTGTGCTCAGACACAGCTGTGACAGAGTTTCTGACGTGAGATGAT 480  
Db 421 GCAAGAGAGGTTTGTGCTCAGACACAGCTGTGACAGAGTTTCTGACGTGAGATGAT 480  
Qy 481 GTGATTAATGCTTCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 540  
Db 481 GTGATTAATGCTTCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 540  
Qy 541 CTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 541 CTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 601 AATAAGTGTGAAAAAGTTCTTTGGACCTTGGATAGCTGTGCAATGAGAGATTAAT 660  
 DB 601 AATAAGTGTGAAAAAGTTCTTTGGACCTTGGATAGCTGTGCAATGAGAGATTAAT 660  
 QY 661 GAACCAACAGAGCAGTGGCATTTGTCAGAGTGTGAGAACGACCTGTGCTCAGATTTGG 720  
 DB 661 GAACCAACAGAGCAGTGGCATTTGTCAGAGTGTGAGAACGACCTGTGCTCAGATTTGG 720  
 QY 721 TATTAACCTGACCTCTTACTAGCTCATGAGAGACAAAAGAGCTTAGGCGAGTTAACTG 780  
 DB 721 TATTAACCTGACCTCTTACTAGCTCATGAGAGACAAAAGAGCTTAGGCGAGTTAACTG 780  
 QY 781 CATTAAGATTTGGCTGAAGTTTATAGAAAGATCTTACAGATAGAGGGCGCATCTGCAAT 840  
 DB 781 CATTAAGATTTGGCTGAAGTTTATAGAAAGATCTTACAGATAGAGGGCGCATCTGCAAT 840  
 QY 841 CCTTGTGTGATGATTTATGAGAGAGGTTTGGTGGAGAGTGAAGATTAATGA 900  
 DB 841 CCTTGTGTGATGATTTATGAGAGAGGTTTGGTGGAGAGTGAAGATTAATGA 900  
 QY 901 ATTGTGTGCTCCAAATGTCATCATGATTAAGCTGAGATTAAGACATTAACGAT 960  
 DB 901 ATTGTGTGCTCCAAATGTCATCATGATTAAGCTGAGATTAAGACATTAACGAT 960  
 QY 961 AAGTGTGCTGGGAGTGGGCAACAGAGCTGCTGGAATTAAGTTCGAGAGATTAAGGCTCTT 1020  
 DB 961 AAGTGTGCTGGGAGTGGGCAACAGAGCTGCTGGAATTAAGTTCGAGAGATTAAGGCTCTT 1020  
 QY 1021 AAGAGCAGCCATTCCTATGATGTCACAGGAGCATGCTGGAGTGAATGTTGATTTGAG 1080  
 DB 1021 AAGAGCAGCCATTCCTATGATGTCACAGGAGCATGCTGGAGTGAATGTTGATTTGAG 1080  
 QY 1081 AAGAGTGCACCTGCTGCTATTTGAGAGGCGCAACGCTTCCACCGGAGTTAGCTGATGGG 1140  
 DB 1081 AAGAGTGCACCTGCTGCTATTTGAGAGGCGCAACGCTTCCACCGGAGTTAGCTGATGGG 1140  
 QY 1141 TTAGATAGAAATTCCTGGGGTCAAGAGCCAGATGTTTTCGAGAGTGTGGCCAACTG 1200  
 DB 1141 TTAGATAGAAATTCCTGGGGTCAAGAGCCAGATGTTTTCGAGAGTGTGGCCAACTG 1200  
 QY 1201 TATGCTCTTCTGCAACGAGAGCAATGCTGACATATTCATCACTCTCAAGGCAAA 1260  
 DB 1201 TATGCTCTTCTGCAACGAGAGCAATGCTGACATATTCATCACTCTCAAGGCAAA 1260  
 QY 1261 ACAAGGCTGAATTCGAGTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAGGCGAG 1320  
 DB 1261 ACAAGGCTGAATTCGAGTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAGGCGAG 1320  
 QY 1321 ATCTCTGAGAGCAATAGAGAGCTGACTCTTAAAGACAAGCTCCAAAACGAGCAAG 1380  
 DB 1321 ATCTCTGAGAGCAATAGAGAGCTGACTCTTAAAGACAAGCTCCAAAACGAGCAAG 1380  
 QY 1381 CAGCCCAAGGCTTGAAGATCTCTGAAATTTATAGAGAGAGAGCTGCTGAGAACTGCA 1440  
 DB 1381 CAGCCCAAGGCTTGAAGATCTCTGAAATTTATAGAGAGAGAGCTGCTGAGAACTGCA 1440  
 QY 1441 GAGGATTAATCGATCTGAGAGACAGAACTAAGATCAGCATGAACAGACAGGAAAGAG 1500  
 DB 1441 GAGGATTAATCGATCTGAGAGACAGAACTAAGATCAGCATGAACAGACAGGAAAGAG 1500  
 QY 1501 ATGGAATGCAACACAGAGGTTTTCATGAGTCAATCAATCAATCAATCAATCAATCAATCA 1560  
 DB 1501 ATGGAATGCAACACAGAGGTTTTCATGAGTCAATCAATCAATCAATCAATCAATCAATCA 1560  
 QY 1561 GCAAAAGAGAGAAATTTGAGATGTTGACAGACAGAAAGTCCCAAGGTTGTTGGCCAG 1620  
 DB 1561 GCAAAAGAGAGAAATTTGAGATGTTGACAGACAGAAAGTCCCAAGGTTGTTGGCCAG 1620  
 QY 1621 CAGACAGAGAACTTAATCTCTTACCAATGACGATTTGCCGAAAGAGCTGAGAGAGTG 1680  
 DB 1621 CAGACAGAGAACTTAATCTCTTACCAATGACGATTTGCCGAAAGAGCTGAGAGAGTG 1680

QY 1681 TCAAGCTTCATGAGTTTCAAGAGAAAGATGAGAGATTTGTTGGAAGAGAGGAGATG 1740  
 DB 1681 TCAAGCTTCATGAGTTTCAAGAGAAAGATGAGAGATTTGTTGGAAGAGAGGAGATG 1740  
 QY 1741 CTGATTAAGATCAAGAGAAAGATGGAAGACATGAAGAAAGGATCAGAGAGATTA 1800  
 DB 1741 CTGATTAAGATCAAGAGAAAGATGGAAGACATGAAGAAAGGATCAGAGAGATTA 1800  
 QY 1801 TTTGATCTGGAAGAAATTTGATGAGGCTTTGGAAGAGCTCATGATTAAGATGGCCTT 1860  
 DB 1801 TTTGATCTGGAAGAAATTTGATGAGGCTTTGGAAGAGCTCATGATTAAGATGGCCTT 1860  
 QY 1861 CACATGAAGATGATGA 1878  
 DB 1861 CACATGAAGATGATGA 1878  
 RESULT 3  
 BT002944  
 LOCUS  
 DEFINITION  
 Arabidopsis thaliana clone RAFL14-93-K05 (R20243) unknown protein  
 (At5g23570) mRNA, complete cds.  
 ACCESSION  
 BT002944  
 VERSION  
 BT002944.1 GI:27754622  
 KEYWORDS  
 FLI CDNA.  
 SOURCE  
 Arabidopsis thaliana (chale cress)  
 ORGANISM  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE  
 1 (bases 1 to 2162)  
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
 Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wong,C., Wu,H.C.,  
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,  
 Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Arabidopsis Full Length cDNA Clones  
 Unpublished  
 2 (bases 1 to 2162)  
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,  
 Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Direct Submission  
 TITLE  
 Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan  
 Street, Albany, CA 94710, USA  
 JOURNAL  
 RIKEN Genomic Sciences Center (GSC) members carried out the  
 collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN  
 Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,  
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
 Hayashizaki,Y. and Shinozaki,K.  
 COMMENT  
 The Salk, Stanford, PGEC (SSP) Consortium members carried out the  
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M.,  
 Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S.,  
 Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G.,  
 Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,  
 Palm,C.J., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to  
 this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)  
 contributed equally to this work as PI.  
 Annotation based on July 2002 version of the Arabidopsis genome  
 submitted to Genbank.  
 FEATURES  
 source  
 1..2162  
 Location/Qualifiers

```
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/ecotype="Raf14-93-K05 (R20243)"
/clone="Columbia"
/notes="This clone is in a modified pBluescript vector2
(lambda PS) as a BamHI/XhoI insert."
1..2162
/gene="At5g23570"
1..91
/gene="At5g23570"
92..1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AA022157.1"
/db_xref="GI:27754623"
/translation="MSSRAGPMSKKNVGGYRPEVQLVGLAGTRLASQDDGGEW
EYISKNNKPKGNTSGKTWVNSNSNPPRMGGQOOGRGNSVSGRGNVSGRNGRG
IOANISGRGRALSRTDNNFVAPPPYRPPLEGGNMQARGSAQHTAQQEPDVEDD
VDNASEENDSDALDSDDDLASDDYSDVSQSHSGSRKQNTMFKKPFESLDSLEQ
INEPQRMHPACQNGPGAIDWYNLHPILAHARTGARVYKDLRELAELKDLQMRG
ASVIPCGEIYQMKGLGEBDEKDEIWPMPVIIMNTRLDKDNKMLGNGNLEIYE
DYELALRAHSYGPQGHGMSVLMFESSATGYLEABRLRELAEMGLDIAMGOKRM
PSGGVROLYGLATRODLDIENHOSGKTRLEKPELKSVMYVXELROI SEDNOIATY
FNKLSKQNKAKVLEBSLEINSEKLRTRAEENRIYRQTKVQHEONREMANHDFE
MDSIKQIHERRDAKEENFEMLDQOERAKVVGQOQNIINSSNDCKRKEVSSPIEF
OEKEMEFEVEREMLIKDEKEMDKKHBEIFDLEKEFDEALEOLMYKRLGHNEED"
1970..2162
/gene="At5g23570"
2146
/gene="At5g23570"
/notes="not present in genomic sequence"

ORIGIN

Query Match 100.0%; Score 1878; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAAAAGAAAGCTTCAGGGTGTATAGGCTT 60
DB 92 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAAAAGAAAGCTTCAGGGTGTATAGGCTT 151
QY 61 GAGTTGAAACAGTTGGTTCAGGTTTGGCAGGACGACGACTGCGCTTCTTCAAGATGAT 120
DB 152 GAGTTGAAACAGTTGGTTCAGGTTTGGCAGGACGACGACTGCGCTTCTTCAAGATGAT 211
QY 121 GGAGGAGAGTGGAGAGTCAATTCCTCAAGAAAGAAACAAACCAAGAAACCTTCTGGA 180
DB 212 GGAGGAGAGTGGAGAGTCAATTCCTCAAGAAAGAAACAAACCAAGAAACCTTCTGGA 271
QY 181 AAAAATGGGTTTCTCAGAAATCGAATCTCTCTAGAGCTTGGGGTGTGTGACGACAAAGG 240
DB 272 AAAAATGGGTTTCTCAGAAATCGAATCTCTCTAGAGCTTGGGGTGTGTGACGACAAAGG 331
QY 241 AGAGGTAGCAACGATCTGGGAGAGGAAACAAATGATCCGGGAGAGGTAAACGCAATGCT 300
DB 332 AGAGGTAGCAACGATCTGGGAGAGGAAACAAATGATCCGGGAGAGGTAAACGCAATGCT 391
QY 301 CGGGGCAATTCAGCTAACATATCTGTCGGGGACGAGCGTTGAGCAAGAAAGTATGATTAAC 360
DB 392 CGGGGCAATTCAGCTAACATATCTGTCGGGGACGAGCGTTGAGCAAGAAAGTATGATTAAC 451
QY 361 AACTTTGTGACACCCCACTGATCTGCGCTCTCTTTGGAAGAGAGATGATTTGGCAG 420
DB 452 AACTTTGTGACACCCCACTGATCTGCGCTCTCTTTGGAAGAGAGATGATTTGGCAG 511
QY 421 GCAAGAGAGAGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTTGAAGTGAAGATGAT 480
DB 512 GCAAGAGAGAGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTTGAAGTGAAGATGAT 571
```

```
QY 481 GTGATTAATGCTTCTGAGGAAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 540
DB 572 GTGATTAATGCTTCTGAGGAAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 631
QY 541 CTTCGAAAGTATGATTAATGATCTCGATGTGACTCAAAAGAGCCATGATCACGAAAGCAG 600
DB 632 CTTCGAAAGTATGATTAATGATCTCGATGTGACTCAAAAGAGCCATGATCACGAAAGCAG 691
QY 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGATTAAGTTTCCATTCAGACAGATTAAT 660
DB 692 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGATTAAGTTTCCATTCAGACAGATTAAT 751
QY 661 GAACCAAGAGGACGATGGGACATGTCAGCTGTCAAGACGACCTGTGCCATTCATTTGG 720
DB 752 GAACCAAGAGGACGATGGGACATGTCAGCTGTCAAGACGACCTGTGCCATTCATTTGG 811
QY 721 TATTAACCTGCAACCTCTACTAGCTCATGCGAGGACAAAAGAGCTAAGCGAGTTAACTTC 780
DB 812 TATTAACCTGCAACCTCTACTAGCTCATGCGAGGACAAAAGAGCTAAGCGAGTTAACTTC 871
QY 781 CATAGGAATTGGCTGAAGTTTATAGAAAAGATCTTACAGATGAGAGCGCATCTGTCAAT 840
DB 872 CATAGGAATTGGCTGAAGTTTATAGAAAAGATCTTACAGATGAGAGCGCATCTGTCAAT 931
QY 841 CCTTGCGTGAGATTTATGGGCAGTGGAAAGGTTTGGGATGAGATGAAGAAAGATTATGAA 900
DB 932 CCTTGCGTGAGATTTATGGGCAGTGGAAAGGTTTGGGATGAGATGAAGAAAGATTATGAA 991
QY 901 ATTTGTCTGCGCTCCAATGCTCATCATGAAATACATGATGATGATGAGACATTAACGAT 960
DB 992 ATTTGTCTGCGCTCCAATGCTCATCATGAAATACATGATGATGATGAGACATTAACGAT 1051
QY 961 AAGTGGCTTGGCATGGGCAACCAAGAGCTGTGGAATCTTTCAGCAAGATGAGGCTCTT 1020
DB 1052 AAGTGGCTTGGCATGGGCAACCAAGAGCTGTGGAATCTTTCAGCAAGATGAGGCTCTT 1111
QY 1021 AGAGCAGGCAATTCCTATGATGTCACAGGCGCATGCGGATGATGTTTGTATGTTTGG 1080
DB 1112 AGAGCAGGCAATTCCTATGATGTCACAGGCGCATGCGGATGATGTTTGTATGTTTGG 1171
QY 1081 AGCAGTGCACATGAGTATTTGGAGGCGCAACGCTTCCACCGGAGTTAGCTGATGAGG 1140
DB 1172 AGCAGTGCACATGAGTATTTGGAGGCGCAACGCTTCCACCGGAGTTAGCTGATGAGG 1231
QY 1141 TTAGATGAATTCCTGCGGCTGAGAAAGCGCAATGTTTCTGAGAGTGTTCGCCAATCTG 1200
DB 1232 TTAGATGAATTCCTGCGGCTGAGAAAGCGCAATGTTTCTGAGAGTGTTCGCCAATCTG 1291
QY 1201 TATGAGCTTCTTGAAGAGAGCAAGATCGACATTAATCAATCAACCTCTCAAGGCANA 1260
DB 1292 TATGAGCTTCTTGAAGAGAGCAAGATCGACATTAATCAATCAACCTCTCAAGGCANA 1351
QY 1261 ACAAGGCTGAATTCGATTTGAATCATCAAGAGATGTTGTAAGAGGCTGAGGCGAG 1320
DB 1352 ACAAGGCTGAATTCGATTTGAATCATCAAGAGATGTTGTAAGAGGCTGAGGCGAG 1411
QY 1321 ATCTTGAGGACATTCAGACGCTGAATCTTTAAGAACACAGCTTCAAAACAGAACAG 1380
DB 1412 ATCTTGAGGACATTCAGACGCTGAATCTTTAAGAACACAGCTTCAAAACAGAACAG 1471
QY 1381 CAGGCAAGAGTCTTGAAGAACTCTGGAATTTAAGAGGAGAAAGCTGCTGAAGCTGCA 1440
DB 1472 CAGGCAAGAGTCTTGAAGAACTCTGGAATTTAAGAGGAGAAAGCTGCTGAAGCTGCA 1531
QY 1441 GAGGATTAATCGGATCGTGAAGACAGAGAACTAAGATGACATGAACAGAACGAGAGAG 1500
DB 1532 GAGGATTAATCGGATCGTGAAGACAGAGAACTAAGATGACATGAACAGAACGAGAGAG 1591
QY 1501 ATGATGCAACAGACAGAGTTTTCATGATTAATCAACAGATCAATGAAGAAAGAGAC 1560
DB 1592 ATGATGCAACAGACAGAGTTTTCATGATTAATCAACAGATCAATGAAGAAAGAGAC 1651
```



QY 1561 GCAAGAGAGAAATTTTCAGATGTTGAGCAGCAGAGAAAGCTGCAAGTTGTTGGCCAG 1620  
| | | | |  
Db 1652 GCAAGAGAGAAATTTTCAGATGTTGAGCAGCAGAGAAAGCTGCAAGTTGTTGGCCAG 1711  
| | | | |  
QY 1621 CAGCAGCAGAACTTAAATCCCTCTAGCAATGACATGTCGCAAGAGAGCTGAGAAAGTG 1680  
| | | | |  
Db 1712 CAGCAGCAGAACTTAAATCCCTCTAGCAATGACATGTCGCAAGAGAGCTGAGAAAGTG 1771  
| | | | |  
QY 1681 TCAAGCTTCATGAGTTTCAAGAGAAAGAGATGAGAGAGTTTGTGAAAGAGAGAGATG 1740  
| | | | |  
Db 1772 TCAAGCTTCATGAGTTTCAAGAGAAAGAGATGAGAGAGTTTGTGAAAGAGAGAGATG 1831  
| | | | |  
QY 1741 CTGATTAAGATCAAGAGAAAGATGAGAGCAATGAAAGAGAGCATCAGAGAGATA 1800  
| | | | |  
Db 1832 CTGATTAAGATCAAGAGAAAGATGAGAGCAATGAAAGAGAGCATCAGAGAGATA 1891  
| | | | |  
QY 1801 TTTGATCTGAGAGAAATTTGATGAGCTTTGAAACAGCTCATGTACAGATGCTT 1860  
| | | | |  
Db 1892 TTTGATCTGAGAGAAATTTGATGAGCTTTGAAACAGCTCATGTACAGATGCTT 1951  
| | | | |  
QY 1861 CACATGAGAGATGATGA 1878  
| | | | |  
Db 1952 CACATGAGAGATGATGA 1969  
| | | | |

RESULT 4  
AF239719 2254 bp DNA linear PLN 02-JUN-2000  
LOCUS Arabidopsis thaliana SGS3 gene, complete cds.  
DEFINITION AF239719  
ACCESSION AF239719.1 GI:8164029  
KEYWORDS

ORGANISM Arabidopsis thaliana (thale cress)  
SOURCE Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Eudicotyledons; Magnoliophyta; Eudicotyledons; core eudicots;  
Spermatophyta; Brassicales; Brassicaceae; Arabidopsids.  
rosids; eudicots II; Arabidopsids.  
1 (bases 1 to 2254)  
Mourrain, P., Becclin, C., Elmayan, T., Feuerbach, F., Godon, C.,  
Morel, J.B., Jonette, D., Lacombe, A.M., Nikic, S., Picault, N.,  
Remoue, K., Santal, M., Vo, T.A. and Vaucheret, H.  
Arabidopsis SGS3 and SGS3 genes are required for  
posttranscriptional gene silencing and natural virus resistance  
Cell 101 (5), 533-542 (2000)

REFERENCE AUTHORS  
TITLE  
JOURNAL MEDLINE  
PUBMED  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .2254  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="V"  
join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)  
/product="SGS3"  
join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)  
/codon\_start=1  
/product="SGS3"  
/protein\_id="AA073960.1"  
/db\_xref="GI:8164030"  
/translation="WSSRAQPMGKQKVGQVPEVQLVQGLAGTTLASODDGEW  
EVIISKNNKNGTSGKTWVSQNSNPPRAQGGQOQSGSVSGKNNVSGKNGRG  
IOANISGRGALSRKYDNFNVAPEVSRPELEGKNNQARCSAQTAAQDFPDVEDD  
VDNASEEDNDLDDSDLDSDYSDVSQSHSRKNNKFKFRSGLDLSIEDO  
INBPORNDHCAQNGGALDWDYVILHLHARTKGRVYKRLAEVLEKDLQNRG  
ASVPCGRTYCGKGLGDEDEYIVPMTYIMTSLDKDDNDKMLGKNDLETF  
DKVEALRNHSYCGQHRGMSVIMFESATGYLAEHLHELMLGMDIRIAMQKSM  
FSGVROLVYGLATKDDLTIPNHSQKTKLKEPLASVGMVVKELRQISDNOQLNY

ORIGIN  
Query Match 51.3%; Score 963; DB 8; Length 2254;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGGCTGCTGCAATGCTTAAGAAAGAAACCTTCAGGGTGTATAGCCT 60  
| | | | |  
Db 1 ATGAGTTCTAGGGCTGCTGCAATGCTTAAGAAAGAAACCTTCAGGGTGTATAGCCT 60  
| | | | |  
QY 61 GAGTTTGAACATTTGCTTCAAGGTTTGGCAGAGAGAGACTGGCTTCTTCAAGATGAT 120  
| | | | |  
Db 61 GAGTTTGAACATTTGCTTCAAGGTTTGGCAGAGAGAGACTGGCTTCTTCAAGATGAT 120  
| | | | |  
QY 121 GAGAGAGTGGAGAGTCAATTTCAAGAAAGCAAGAAACCAACAGAAACACTTGGGA 180  
| | | | |  
Db 121 GAGAGAGTGGAGAGTCAATTTCAAGAAAGCAAGAAACCAACAGAAACACTTGGGA 180  
| | | | |  
QY 181 AAACTTGGGTTCTGCAATTCGAATCCTCTAGAGCTGGGTTGTCAGCAGCAAGG 240  
| | | | |  
Db 181 AAACTTGGGTTCTGCAATTCGAATCCTCTAGAGCTGGGTTGTCAGCAGCAAGG 240  
| | | | |  
QY 241 AGAGTGAACAAGTATCTGGAGAGAAACAATGATCCGGAGAGTTAACCGCAATGAT 300  
| | | | |  
Db 241 AGAGTGAACAAGTATCTGGAGAGAAACAATGATCCGGAGAGTTAACCGCAATGAT 300  
| | | | |  
QY 301 CGGGGCAATCAAGCTTAACATATCGTGGGAGAGAGCTTGAAGAGAAAGTATGATAC 360  
| | | | |  
Db 301 CGGGGCAATCAAGCTTAACATATCGTGGGAGAGAGCTTGAAGAGAAAGTATGATAC 360  
| | | | |  
QY 361 AACTTTGGGCAACCCCAACCTGATATCTGCGCTCTTTGAGAGAGAGATGGAGCAG 420  
| | | | |  
Db 361 AACTTTGGGCAACCCCAACCTGATATCTGCGCTCTTTGAGAGAGAGATGGAGCAG 420  
| | | | |  
QY 421 GCAAGAGAGGTTTCTGCTGACACACAGCTGTCAGAGAGTTTCTGACGTGAGATGAT 480  
| | | | |  
Db 421 GCAAGAGAGGTTTCTGCTGACACACAGCTGTCAGAGAGTTTCTGACGTGAGATGAT 480  
| | | | |  
QY 481 GTGATTAATGCTTCTGAGAGAGAGATGATCCGATCTTTGATGATGATGATGATGAT 540  
| | | | |  
Db 481 GTGATTAATGCTTCTGAGAGAGAGATGATCCGATCTTTGATGATGATGATGATGAT 540  
| | | | |  
QY 541 CTTGCAAGTATGATTAATGATCTCGGATGATGATGATGATGATGATGATGATGATGAT 600  
| | | | |  
Db 541 CTTGCAAGTATGATTAATGATCTCGGATGATGATGATGATGATGATGATGATGATGAT 600  
| | | | |  
QY 601 AATAAGTGTTCAGAAAGATCTTTGGCAGCTTGTGATGATGATGATGATGATGATGAT 660  
| | | | |  
Db 601 AATAAGTGTTCAGAAAGATCTTTGGCAGCTTGTGATGATGATGATGATGATGATGAT 660  
| | | | |  
QY 661 GAACCAAGAGGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
| | | | |  
Db 661 GAACCAAGAGGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
| | | | |  
QY 721 TATTAACCTGACCTTCTAGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 780  
| | | | |  
Db 721 TATTAACCTGACCTTCTAGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 780  
| | | | |  
QY 781 CATTAAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
| | | | |  
Db 781 CATTAAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
| | | | |  
QY 841 CTTTGTGATGATTTATGAGGAGTGGAGAGGTTTGGTGTGAGATGATGATGATGATGAT 900  
| | | | |  
Db 841 CTTTGTGATGATTTATGAGGAGTGGAGAGGTTTGGTGTGAGATGATGATGATGATGAT 900  
| | | | |  
QY 901 ATTGTCTGCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
| | | | |  
Db 901 ATTGTCTGCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
| | | | |



```

QY      961 AAG 963
Db      961 AAG 963

RESULT 5
AX078760      3275 bp      DNA      linear      PAT 22-FRB-2001
LOCUS      Sequence 1 from Patent WO0105951.
DEFINITION
ACCESSION      AX078760
VERSION      AX078760.1 GI:13158379
KEYWORDS
SOURCE
ORGANISM      Arabidopsis thaliana (thale cress)
              Arabidopsis thaliana
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
1      Beclin, C., Elmayer, T. and Vaucheret, H.
    Novel sgas plant gene and use thereof
    Patent: WO 0105951-A 1 25-JAN-2001;
    AVENTIS CROPS SCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
    AGRONOMIQUE (FR)

FEATURES
    source
        1..3275
        /organism="Arabidopsis thaliana"
        /mol_type="unassigned DNA"
        /db_xref="taxon:3702"
        complement(2926..2952)
        primer_bind
            primer_bind
                /note="p356AD"
                complement(2926..2952)
                /note="p356Y"

ORIGIN

Query Match      51.3%; Score 963; DB 6; Length 3275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGAGTTCTAGAGGCTGCTCCATGTTTAAAGAAAAGAAAGCTTCAGAGGCTGTTATAGGCTT 60
Db      636 ATGAGTCTTAGAGGCTGCTCCATGTTTAAAGAAAAGAAAGCTTCAGAGGCTGTTATAGGCTT 755

QY      61 GAGGTTGAACAGTTGGTTCAAGGTTTGAAGGAGAGAGAGCTTCTTCAACAAGTAT 120
Db      756 GAGGTTGAACAGTTGGTTCAAGGTTTGAAGGAGAGAGAGCTTCTTCAACAAGTAT 815

QY      121 GGAGAGAGAGTGGAGAGTCTTTCAGAGAAACAAGAACAAACAGAGAAACCTTCTGGA 180
Db      816 GGAGAGAGAGTGGAGAGTCTTTCAGAGAAACAAGAACAAACAGAGAAACCTTCTGGA 875

QY      181 AAAAAGTGGGTTTCTCAGATTCGATTCCTCTAGAGCTTGGGCTGTCAGACAGCAAGG 240
Db      876 AAAAAGTGGGTTTCTCAGATTCGATTCCTCTAGAGCTTGGGCTGTCAGACAGCAAGG 935

QY      241 AAGAGTGAACAGTATCTGGAGAGAGAAACATGTTCCGGAGAGAGTAAACGCAATGCT 300
Db      936 AAGAGTGAACAGTATCTGGAGAGAGAAACATGTTCCGGAGAGAGTAAACGCAATGCT 995

QY      301 CGGGGCAATTCAGCTTAACATATCTGGTCCGGAGAGAGAGCTTGAAGCAAGATATGATTAAC 360
Db      996 CGGGGCAATTCAGCTTAACATATCTGGTCCGGAGAGAGAGCTTGAAGCAAGATATGATTAAC 1055

QY      361 AACTTTGTGGACACCCCACTGATCTGCCTCTCTTGAAGAGAGATGAAATGGCAG 420
Db      1056 AACTTTGTGGACACCCCACTGATCTGCCTCTCTTGAAGAGAGATGAAATGGCAG 1115

QY      421 GGAAGAGAGAGTTCGTCAGACACAGCTGAGAGAGTTCCTGACGTCAGAGATGAT 480
Db      1116 GGAAGAGAGAGTTCGTCAGACACAGCTGAGAGAGTTCCTGACGTCAGAGATGAT 1175

QY      481 GTGATTAATGCTTCGAGAGAGAGATGATTCGATGCTTGGATGATTCGATGACGAC 540

```

```

Db      1176 GTGATTAATGCTTCGAGAGAGAGATGATTCGATGCTTGGATGATTCGATGACGAC 1235
QY      541 CTGCAAGAGATGATTAATGATCGATGATGATCAAAAAGAGCAGATGATCAGAAAAGCAG 600
Db      1236 CTGCAAGAGATGATTAATGATCGATGATGATCAAAAAGAGCAGATGATCAGAAAAGCAG 1295
QY      601 AATAAGTGGTTCAAAAAGTCTTGGACGCTTGGATGATGCTTGTGATGAGCAGATTAAT 660
Db      1296 AATAAGTGGTTCAAAAAGTCTTGGACGCTTGGATGATGCTTGTGATGAGCAGATTAAT 1355
QY      661 GAACCAAGAGAGAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db      1356 GAACCAAGAGAGAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1415
QY      721 TATTAACCTGACCCCTTACTAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db      1416 TATTAACCTGACCCCTTACTAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1475
QY      781 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAGAG 840
Db      1476 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAGAG 1535
QY      841 CCTTGTGAGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db      1536 CCTTGTGAGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1595
QY      901 ATTGTGCGCCCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db      1596 ATTGTGCGCCCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1655
QY      961 AAG 963
Db      1656 AAG 1658

```

```

RESULT 6
AB025633
LOCUS      Arabidopsis thaliana genomic DNA, chromosome 5, Pl clone:MQM1.
DEFINITION
ACCESSION      AB025633 BA000015
VERSION      AB025633.2 GI:10178221
KEYWORDS
SOURCE
ORGANISM      Arabidopsis thaliana (thale cress)
              Arabidopsis thaliana
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
1      (sites)
    Sato, S., Nakamura, Y., Kaneo, T., Katoh, T., Asanizu, E., Kotani, H.
    and Tabata, S.
    Structural analysis of Arabidopsis thaliana chromosome 5. X.
    Sequence features of the regions of 3,076,755 bp covered by sixty
    P1 and TAC clones
    DNA Res. 7 (1), 31-63 (2000)

JOURNAL
MEDLINE
PUBMED      10718197
REFERENCE
2      (bases 1 to 81365)
    Nakamura, Y.
    Direct Submisison
    Submitted (02-Apr-1999) Yasukazu Nakamura, Kazusa DNA Research
    Institute, Department of Plant Gene Research; 1532-3, Yana,
    Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
    Tel:81-438-52-3935, Fax:81-438-52-3934)
    On Sep 15, 2000 this sequence version replaced gi:1589439.
    Address for correspondence: kaos@kazusa.or.jp
    For the latest information on annotation of this clone, please see
    http://www.kazusa.or.jp/kaos/cgi-bin/aggd.graph.cgi?c=MQM1
    Genes with similarity to proteins in the databases are described in
    'product' or 'note' qualifiers. Genes that have no significant
    protein similarity are described as 'unknown protein'.
    The software programs used to predict genes include: Grail
    (Informatics Group, Oak Ridge National Laboratory,
    http://comblab.ornl.gov/Grail-1.3/),

```

GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),  
Necenz2 (S.M. Hebsgaard, et al., CBS, Technical University of  
Denmark, <http://www.cbs.dtu.dk/services/Necenz2/>) and  
SplicePredictor (Volker Brendel, Stanford University,  
<http://gremlii.zool.iastate.edu/cgi-bin/sp.cgi>).  
Genes encoding tRNAs are predicted by tRNAscan-SE  
(Sean Eddy, Washington University School of Medicine, St. Louis,  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).  
This sequence may not be the entire insert of this clone. It may be  
shorter because we remove overlaps between neighboring submissions.  
The 5' clone is K19M3 and the 3' clone is MRO11.

FEATURES  
source

Location/Qualifiers

1. 81365

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/strain="Columbia"

/db\_xref="taxon:3702"

/chromosome="5"

/clone\_1ib="M1cui P1"

join(3184..3403,3474..3595,3681..3723,3777..3898,

4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)

/note="gb|AA01580.1

gene\_id:MOM1.24

similar to unknown protein"

/evidence=not experimental

/codon\_start=1

/protein\_id="BA097251.1"

/db\_xref="GI:8809710"

/translation="MHVQPTSDVSEPAHKVDTEPPKAVQLEPISTKVVDNAQNS

YQAFDDPSSNSPLISVFEPPSSFSFGQDDPFGIEDIQISEPYGHEIACG

YSINGTSCNFEVCHLEDGSVNYIDESRERTSNAAGQPMVLTADVDYLAIEV

QGLDNRNRKGFAPGKGYDLVLVDVHTLFSADCFSEPFMMFAICYEELISG

ACEDDMOSNIKTHTGHSYKSLAVGVNIMWATLSIRBGVSIKICISDLTAA

KESASTVTIIPFGQAEVITISDSSEHSRLDNGSPDLIGSRDELVTLRPIKRW

FKALNSFYHFCSCSKSRGNL"

join(6317..6428,6508..6629,6714..6756,6841..6932,

7121..7169,7396..7466,7550..7663,7785..7882,8112..8181)

/note="gb|AA01580.1

gene\_id:MOM1.23

similar to unknown protein"

/evidence=not experimental

/codon\_start=1

/protein\_id="BA097250.1"

/db\_xref="GI:8809709"

/translation="MQPGTSYSEPALNADDETPISKTUVNDTKNIPEYDDGAPL

PVLBNQISGEPPYGHEDACGYSINGTSCNFEVCHLEDGSVNYIDGAKKPNLTAT

ADVDGLCAIEVQPLDNRKGLVVFANDRKIAKCHPEMOSNIDKTLHTGHSYK

SLAIGFVHMEATLSIEREGYTIKCNNDLTITKFSASTAVKIPFEKPAELVIIISD

GSEHCIRVNEPDIRLCKEERVLALVYVVFQKTL"

complement(join(8642..8705,8794..8863,8978..9387,

9584..10347))

/note="gene\_id:MOM1.22

unknown protein"

/codon\_start=1

/evidence=not experimental

/protein\_id="BA097249.1"

/db\_xref="GI:8809708"

/translation="MSWKGKSSGWTAPDLKROKQGLSEYBGPFPVSVTSVNASF

GVGRRLRNHSEPSKSSVLLPSPRFPALTENKOCNGERGCGCKRCKDTLSLPNS

HDLAFLTKEMNSWADNLIIRDVLSTEDFEMALFLKGMVSGCKDEDEPSKIEGY

SDNRSEYRTFEKTVTSVKKAASTFEDAGVLENSDGSFLVNAADNKPFDPI

SLDSIIORLOSTPIEBWEEDLVLSHRKDLKVMRSASNSRAQNAFOGYDHASA

KQHSDEKEDMLAAEKLNAEAKKIGITGNKNDIKLIDHGLHATEVQALQERLOM

IGHFTVNSVSPNKRGRKNAALRSASQSPFGALDEBGNHCORTSRELRNSLOVITG

ICKHSRQASLPLVATPFEDNRKRFDETRPGVITVRPFRHS"

complement(12417..13424)

## CDS

/note="contains similarity to unknown protein

gb|AA097018.1

gene\_id:MOM1.21"

/codon\_start=1

/evidence=not experimental

/protein\_id="BA097248.1"

## CDS

/db\_xref="GI:8809707"

/translation="MAIDSQPNQKLTLEPKTRIALPTVISTMTDANARBDGTINRFLR

LEDFRAPNPKNVINSDEEVQOSRDLNRLYTPHVSQDKIPIVVEFHGGAFSL

PNAYPVNDCRPARKILPAAVYIVNRVAPRRRPAVDGDFDLKYLENHGSLIPA

NADLSRCFRPADSDAGNIAHNVAIRICRPPSSFLAVYLIGLISGPFGGERTAE

KQVGAPLVSFDRDTCMKANLNRDRAVVGGBNAVIDISGLDIPETMVVYAGDPL

KDWORSYEWKLCQKATTLIEYPMFAFYIPELPAPAGOLIMIKIKOFVERVASTS

A"

join(14922..14981,15072..15167,15563..15666,15755..15891,

16096..16178)

/note="gene\_id:MOM1.20"

/codon\_start=1

/evidence=not experimental

/product="50S ribosomal protein L24"

/protein\_id="BA097247.1"

/db\_xref="GI:8809706"

/translation="MGKKAARKLIRHKILRGDNVMIIRGKDKETGTXKVRISQR

VIVEGKILKHKGPDHGGIFTEVPELHASNQVQVDPVTRGCKVGRVLEDTGK

VRVARGTSGSIIIPRELIRATPRPTTGPDTPAEFWEGTVDKTKGMPDL"

join(16823..16952,17143..17320,17412..17493,17675..17851,

17943..18110,18199..18390)

/note="gene\_id:MOM1.19"

/codon\_start=1

/evidence=not experimental

/product="26S proteasome, non-ATPase regulatory subunit"

/protein\_id="BA097246.1"

/db\_xref="GI:8809705"

/translation="MERLRIIRGAGGIGHASPDSPTLDSBOVYISLALKMLKHG

RAGVMEVGMGLGFEVDERTVRYVDVYAMQSGVSEVAVDHYFQNMIDMLKQTS

RPEVAVGMYHSHPGFQGLSGVDINTQOSFALNORAVAVVDPIOSYKGVADIAFR

SINPTQIMLQGEPRQTSNLHLNKPSTQALIHGNRYYSIAINRYRNELEKMLN

LHKKMTDGLTRFEDTSTKNEQVQMSLAKYKNAVOEBEELSPEKLAIVWGR

ODAKKHLEHYSNLSNIVOTLGTMLDVTVF"

complement(join(18965..19043,19240..19298,19419..19536,

19666..19739,19908..19961,20387..20530))

/note="contains similarity to unknown protein

emb|CAA2897.1

gene\_id:MOM1.18"

/codon\_start=1

/evidence=not experimental

/protein\_id="BA097245.1"

/db\_xref="GI:8809704"

/translation="MDKKNQALIKKMKLVMEVEDEQQADESSLSFMEDLNRCAL

TYKRFYGAICLSAGLTCTTSLMLVFNPVKFGITFTLGMALMGSTAPLIGPQRY

TMMLDPAIVATVATYLAISIIIALFCALYVRNKLITLALIIIEFTLIWYSYIPAR

TMVSKIMQTCDETER"

join(22672..22634,23708..23999,24111..24355,24458..24624,

24715..24925)

/note="emb|CAB2356.1

gene\_id:MOM1.17

similar to unknown protein"

/codon\_start=1

/evidence=not experimental

/protein\_id="BA097244.1"

/db\_xref="GI:8809703"

/translation="MSSRACGPMSEKKNVQGYRPEVEQLVOGLAGTRIASQDDGEM

EVISKKNKPKGNTSGKTWSONSNPPRAGGQOQGRGNSGRNNTVSGNGRG

## CDS

join(25614..25838,26061..26207,26324..26438,26516..26610,

26710..26813,27522..27765,27943..28036,28115..28305,

28385..28604,28688..28855,28944..29092)

/note="gb|AAC97420.1

gene\_id:MOM1.16

strong similarity to unknown protein"

```

/codon_start=1
/evidence=not experimental
/protein_id="BAA97243.1"
/db_xref="GI:8809702"
/translation="MAPRAGOTTAVAEAGADGAPPOOGRGRTISGIVRAVWYF
ASKEPSPOKEMDPTAASQMLNTLPHKESLDMWYFSEOEKFNFGDRALYHEBTN
IPIVAWTPESIRTKSLIYVPSRTLQNSLSVHIFPASGSDIDTDEYQDLMKPSR
THAVATYEPKOKKSKSLISGKDSDESEPEVKYDGPKEEVEVEMSLKMPN
VTINLVDDTFADLVEPTNGNYPTIYFNEFMILRDKFIPIVNEVSELPINLEISF
ISMWKOLFOOVODSFOMORSYSGWLDSESDLKVELEGNYVLLGITMFEVMSHSVF
DPLAFKNDIQFNKSKSMEGLSAKSVLNFICQFIYFLINDNTSMILASSGVC
IEFWKIGKRLIYDVSQWIPRLRPHDEBSVSNKTEKEDAIKFLSVLLVILVGL
SIYSLAVERHSMYSWILSLTSVYMEGFIKMCQPLINYLKSKVAHLPKMOYKFK
LNTIIDLPAFVILKPIILRLSVFRDVIYFLIYLRWVYPDKTRVIEFGGDEET
AEKLIITEKEEDDKTN"
complement(join(29439..29576,29661..29885,29985..30268,
30341..30493,30579..30722,30956..31484))
/note="gene_id:MOM1.15

```

Query Match 51.3%; Score 963; DB 8; Length 81365;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAAAAGAAAGCTTCAAGGTTATAGGCT 60
22672 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAAAAGAAAGCTTCAAGGTTATAGGCT 22731
61 GAGGTTGAACAAGTTGGTTCAAGGTTGGGAGGAGCAAGCTTCTTCAACAAGTAT 120
22732 GAGGTTGAACAAGTTGGTTCAAGGTTGGGAGGAGCAAGCTTCTTCAACAAGTAT 22791
121 GGAGGAGAGTGGAGGATCTATTCGAAGAAAGAAACAACAACCGAGAAACCTTCTGGA 180
22792 GGAGGAGAGTGGAGGATCTATTCGAAGAAAGAAACAACAACCGAGAAACCTTCTGGA 22851
181 AAAAAGTGGGTTTCTCAAGATTCGAATCCTCTTGAAGCTTGGGAGTGTGTCAGCAGCAAGG 240
22852 AAAAAGTGGGTTTCTCAAGATTCGAATCCTCTTGAAGCTTGGGAGTGTGTCAGCAGCAAGG 22911
241 AGAGGTACGACGATTCGAGGAGAGAAACAATGATTCGGGAGAGGATTAACGGCAATGCT 300
22912 AGAGGTACGACGATTCGAGGAGAGAAACAATGATTCGGGAGAGGATTAACGGCAATGCT 22971
301 CGGGGCAATTCAGGATTAACGATTCGAGGAGAGCAAGGCTTGAAGAAAGATATATAC 360
22972 CGGGGCAATTCAGGATTAACGATTCGAGGAGAGCAAGGCTTGAAGAAAGATATATAC 23031
361 AACTTGTGAGCAAGGCTGATTCGAGGAGAGCAAGGCTTGTGGAAGAGAGATGATGAGCAG 420
23032 AACTTGTGAGCAAGGCTGATTCGAGGAGAGCAAGGCTTGTGGAAGAGAGATGATGAGCAG 23091
421 GCAAGAGAGAGTTCTGCTCAGACACAGCTGTGCAAGAGTTCTGTCGATGAGAGATGAT 480
23092 GCAAGAGAGAGTTCTGCTCAGACACAGCTGTGCAAGAGTTCTGTCGATGAGAGATGAT 23151
481 GTGATATGCTTCTGAGGAAGAGATGATTCGAGGATTCGATGATGATGATGATGATGATGAT 540
23152 GTGATATGCTTCTGAGGAAGAGATGATTCGAGGATTCGATGATGATGATGATGATGATGAT 23211
541 CTTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
23212 CTTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 23271
601 AATAAGTGTTCACAAAAGTTCTTGGCAGCTTGAATGCTTGTGATGAGCAGATTAAT 660
23272 AATAAGTGTTCACAAAAGTTCTTGGCAGCTTGAATGCTTGTGATGAGCAGATTAAT 23331
661 GAACACAGAGGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
23332 GAACACAGAGGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 23391
721 TATAACTGCAAGCTTCTACTAGCTATGAGCAGAGCAAAAAGAGCTTAGGCAATTAAGCTC 780

```

```

Db 23392 TATAACTGCAAGCTTCTACTAGCTATGAGCAGAGCAAAAAGAGCTTAGGCAATTAAGCTC 23451
Qy 781 CATAGAGATTTGGCTGATGATTTAGAAAAGATCTACAGATAGAGGCGCATCTGTCAAT 840
Db 23452 CATAGAGATTTGGCTGATGATTTAGAAAAGATCTACAGATAGAGGCGCATCTGTCAAT 23511
Qy 841 CTTGTGTGTGATGATTTATGAGGAGATGAAAGGCTTTGGTGTGAGATGAAAAGATTAATGAA 900
Db 23512 CTTGTGTGTGATGATTTATGAGGAGATGAAAGGCTTTGGTGTGAGATGAAAAGATTAATGAA 23571
Qy 901 ATTGTCTGAGCTTCCCATGATGATCATGATGAAATCTAGCTGATTAAGAGATTAACAT 960
Db 23572 ATTGTCTGAGCTTCCCATGATGATCATGATGAAATCTAGCTGATTAAGAGATTAACAT 23631
Qy 961 AAG 963
Db 23632 AAG 23634

```

RESULT 7  
 ATHS28171  
 LOCUS  
 DEFINITION  
 Arabidopsis thaliana T-DNA flanking sequence, left border, clone 157C08.  
 ACCESSION  
 AJ528171  
 VERSION  
 AJ528171.1 GI:26796431  
 KEYWORDS  
 left border; T-DNA flanking sequence.  
 SOURCE  
 Arabidopsis thaliana  
 ORGANISM  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE  
 1  
 Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Sanson, F., Chauvin, S., Bechthold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepoint, L., Caboche, M. and Lecharny, A.  
 T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites  
 EMD Rep. 3 (12), 1152-1157 (2002)  
 MEDLINE  
 22363535  
 PUBMED  
 12446565  
 REFERENCE  
 2 (bases 1 to 650)  
 Balzergue, S.  
 DIRECT SUBMISSION  
 TITLE  
 Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE  
 JOURNAL  
 Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE  
 COMMENT  
 PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment (s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr/>).  
 FEATURES  
 source  
 1..650  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /cultivar="Wassiljewskij3"  
 /db\_xref="taxon:3702"  
 /clone="157C08"  
 /clone\_1lb="Arabidopsis thaliana T-DNA insertion lines"  
 misc\_feature  
 1..650  
 /note="T-DNA flanking sequence  
 left border"

Query Match 12.7%; Score 239; DB 8; Length 650;  
 Best Local Similarity 99.7%; Pred. No. 9.9e-131;  
 Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



```

misc_feature      /mol_type="genomic DNA"
                  /db_xref="taxon:10116"
                  /clone="CH230-32812"
                  1..1169
                  /note="wgs_end_extension
                  clone_end:5p6"
misc_feature      /note="wgs_end_extension
                  clone_end:5p6"
                  27453..29430
                  /note="wgs_end_extension
                  clone_end:5p6"
                  77608..79432
                  /note="wgs_end_extension
                  clone_end:5p6"
                  99609..100719
                  /note="wgs_end_extension
                  clone_end:5p6"
                  complement(118199..119034)
                  /note="clone_boundary
                  clone_end:5p6
                  site:MboI
                  end_sequence:RXAON49TV"
                  125929..127858
                  /note="wgs_contig"
                  156410..157915
                  /note="wgs_contig"
                  complement(172023..172895)
                  /note="clone_boundary
                  clone_end:T7
                  site:MboI
                  end_sequence:RXAON49TV"
                  173824..175436
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature

```

## ORIGIN

```

Query Match      1.2%; Score 22; DB 2; Length 178928;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 492 TTCTGAGGAGAGATGATTC 513
Db 112510 TTCTGAGGAGAGATGATTC 112489

```

```

RESULT 9
LOCUS AL928719/c 197909 bp DNA linear ROD 28-JAN-2003
DEFINITION Mouse DNA sequence from clone RP23-419G21 on chromosome 2, complete
sequence.
ACCESSION AL928719
VERSION AL928719.6 GI:24527684
KEYWORDS HTG
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 197909)
REFERENCE 1
AUTHORS Leonamornler, D.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Nov 3, 2003 this sequence version replaced gi:24395353.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----

```

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. Unless otherwise noted: all this sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M3 subclone; and the rare assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Mp, WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-419G21 is from the RPI-23 Mouse BAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

## FEATURES

```

source
  1..197909
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="2"
  /clone="RP23-419G21"
  /clone_1fb="RPI-23"

```

## ORIGIN

```

Query Match      1.2%; Score 22; DB 10; Length 197909;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 567 TGTGAGTCAGAAAGCCATGGA 588
Db 34650 TGTGAGTCAGAAAGCCATGGA 34629

```

```

RESULT 10
LOCUS AC078933 214186 bp DNA linear HTG 11-AUG-2000
DEFINITION Mus musculus chromosome 5 clone RP23-201B13 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 17 unordered pieces.
ACCESSION AC078933
VERSION AC078933.1 GI:9795569
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214186)
REFERENCE 1
AUTHORS Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Dietrich, N.L., Eagle, W.O., Gupta, J., Ho, S.-L.,
Huang, M.C., Idol, J., Lee-Jin, S.-Q., Maduro, O.L., Maduro, V.B.,
Maestrian, S.D., McCloskey, J.C., McDowell, J., Ojodu, M.A., Pearson, R.,
Stancitrop, S., Summers, T.J., Thomas, J.W., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,
Wehner, R.D. and Green, E.D.
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214186)
AUTHORS Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2000) NIH Intramural Sequencing Center, 8717
Groveview Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
-----

```

Query Match	1.2%;	Score 22;	DB 2;	Length 214186;
Best Local Similarity	100.0%;	Pred. No. 2.1;		
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Db	99680	TGTGAGTCAAAAGAGCCATGGA	99701	
RESULT 11				
AC079182/c				
LOCUS				
DEFINITION				
AC079182				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
JOURNAL				
TITLE				
COMMENT				



[illegible]



Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Kapahy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kover, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, D., Lorenshaw, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundiada, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, G., Nguyen, N., Norris, S., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poldexter, A., Popovic, D., Prims, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sison, I., Sitter, C.D., Smajic, D., Sneed, A., Sodergren, B., Song, X.-Z., Sorelle, R., Sosa, J., Steinhilber, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, D., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., Von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Unpublished  
Direct Submission  
2 (bases 1 to 236508)

Worley, K.C.  
Direct Submission  
Submitted (17-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 236508)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 23, 2002 this sequence version replaced gi:22772519. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----  
Center project name: GBS  
Center clone name: CH230-11F1

----- Summary Statistics -----  
Assembly program: Phrap; version 0.990329  
Consensus quality: 219730 bases at least Q40  
Consensus quality: 22791 bases at least Q30  
Consensus quality: 224323 bases at least Q20  
Estimated insert size: 225524; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 125318: contig of 125318 bp in length  
\* 125319 125418: gap of unknown length in length  
\* 125419 180275: contig of 54857 bp in length  
\* 180276 236508: gap of unknown length  
\* 180376 236508: contig of 56133 bp in length.

Location/Qualifiers  
1. 236508 "Rattus norvegicus"  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-11F1"

misc\_feature  
1. 1641  
/note="wgs contig"  
misc\_feature  
15976..157691  
/note="wgs contig"  
misc\_feature  
233371..235181  
/note="wgs contig"  
misc\_feature  
235232..236508  
/note="wgs contig"

ORIGIN  
Query Match 1.2%; Score 22; DB 2; Length 236508;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 AACGAGAAACACTTCGAAA 182  
Db 162467 AACGAGAAACACTTCGAAA 162446

RESULT 13  
AC140364 280016 bp DNA linear HTG 05-NOV-2003  
LOCUS Mus musculus chromosome UNK clone RP24-75K5, WORKING DRAFT  
DEFINITION  
SEQUENCE, 8 unordered pieces.  
ACCESSION AC140364.2 GI:38176012  
VERSION HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_FULTOP.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 280016)  
Wilson, R.K.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 280016)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 280016)  
Wilson, R.K.  
Direct Submission  
Submitted (05-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
On Nov 5, 2003 this sequence version replaced gi:28475622.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: submissions@wustl.edu

----- Project Information -----

Center project name: M\_BB0075K05

----- Summary Statistics -----

Sequencing vector: M13; 0%  
Chemistry: dye-primer ET; 0% of reads  
Chemistry: dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 278028 bases at least Q40  
Consensus quality: 278887 bases at least Q30  
Consensus quality: 279492 bases at least Q20

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1202: contig of 1202 bp in length  
\* 1203 1302: gap of unknown length  
\* 1303 3866: contig of 2564 bp in length  
\* 3867 3966: gap of unknown length  
\* 3967 24768: contig of 20802 bp in length  
\* 24769 24868: gap of unknown length  
\* 24869 48433: contig of 23565 bp in length  
\* 48434 48533: gap of unknown length  
\* 48534 85803: contig of 37270 bp in length  
\* 85804 85903: gap of unknown length  
\* 85904 127685: contig of 41782 bp in length  
\* 127686 127785: gap of unknown length  
\* 127786 198320: contig of 70435 bp in length  
\* 198321 280016: gap of unknown length  
\* 198321 280016: contig of 81696 bp in length.

FEATURES

source 1. 280016

/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP24-75K5"

misc\_feature 1. 1202  
/note="assembly\_name:Contig20"  
misc\_feature 1303. 3866  
/note="assembly\_name:Contig22"  
misc\_feature 3967. 24768  
/note="assembly\_name:Contig23"  
misc\_feature 24869. 48433  
/note="assembly\_name:Contig24"  
misc\_feature 48534. 85803  
/note="assembly\_name:Contig25"  
misc\_feature 85904. 127685  
/note="assembly\_name:Contig26"  
misc\_feature 127786. 198320  
/note="assembly\_name:Contig27"  
misc\_feature 198321. 280016  
/note="assembly\_name:Contig28"

ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 280016;  
Best Local Similarity 100.0%; Pred.No. 2.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 567 TGTGAGTCAAAAGACCATGA 568  
|||||  
DB 58633 TGTGAGTCAAAAGACCATGA 58654

RESULT 14  
AC073813 304407 bp DNA linear HTG 29-JUN-2000  
LOCUS AC073813  
DEFINITION Mus musculus clone RP23-69B1, WORKING DRAFT SEQUENCE, 58 unordered

pieces.  
AC073813  
AC073813.1 GI:8810430  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 304407)  
DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 304407)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

Project Information  
Center Project Name: 1763393  
Center clone name: RP23-69B1

----- Summary Statistics -----

Consensus quality: 257140 bases at least Q40  
Consensus quality: 282847 bases at least Q30  
Consensus quality: 288639 bases at least Q20  
Estimated insert size: 272860; agarose-tp estimation  
Estimated coverage: 9.46 in Q20 bases; agarose-tp estimation  
Quality coverage: 8.64 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 58 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1167: contig of 1167 bp in length  
\* 1168 1267: gap of unknown length  
\* 1268 2281: contig of 1014 bp in length  
\* 2282 2381: gap of unknown length  
\* 2382 3767: contig of 1386 bp in length  
\* 3768 3867: gap of unknown length  
\* 3868 5372: contig of 1505 bp in length  
\* 5373 5472: gap of unknown length  
\* 5473 6494: contig of 1022 bp in length  
\* 6495 6594: gap of unknown length  
\* 6595 7909: contig of 1315 bp in length  
\* 7910 8009: gap of unknown length  
\* 8010 9416: contig of 1407 bp in length  
\* 9417 9517: gap of unknown length  
\* 9517 10776: contig of 1260 bp in length  
\* 10777 10876: gap of unknown length  
\* 10877 12279: contig of 1403 bp in length  
\* 12280 12379: gap of unknown length  
\* 12380 13920: contig of 1541 bp in length  
\* 13921 14020: gap of unknown length  
\* 14021 15417: contig of 1397 bp in length  
\* 15418 15517: gap of unknown length  
\* 15518 16897: contig of 1380 bp in length  
\* 16898 16997: gap of unknown length  
\* 16998 18940: contig of 1943 bp in length  
\* 18941 19040: gap of unknown length  
\* 19041 20890: contig of 1850 bp in length  
\* 20891 20990: gap of unknown length  
\* 20991 22965: contig of 1975 bp in length  
\* 22966 23065: gap of unknown length  
\* 23066 24501: contig of 1436 bp in length



gene 125..1510  
/gene="huth"  
/note="synonym: VNG1212G"  
CDS 125..1510  
/gene="huth"  
/note="huth"  
/codon\_start=1  
/transl\_table=11  
/product="histidine ammonia-lyase"  
/protein\_id="AAG19580.1"  
/db\_xref="GI:10580744"  
/translation="MSDTRIDADREBALQANVRSNAGAGSELDTAVALRALLVRLN  
ALAKYSIRERLDVLVGLNVEHPVPPSKSLGASGDLAPLAMSRVLLGEOQAD  
VAGERMPAAELAAADLEPVTLOAKEGALINGTQLTGVALLADHERVRSADTA  
GALTEVEMSTASCAPAIHEVRPHDGOVASRHIRNLJAGSEVLDHHRDCRVODAY  
SIRCLPOVHAGVRLDHLRAVATELNSATDNPVPAGDRVGPRAASGDRPAVVSIG  
NPHGVALARIGYAAALAEALAIISERRDRILNPETOQPHLEFLAPDSGHSGLMT  
FOYTNASLVNDLRSLGQPTLDNASVSGAODEHVSAGAIYFREAVEKATVAGVEL  
LCGAGREFLDPLDLAGTAAAYDLVREVSPPVAGDRALADMAVGDVRAAGVED  
AVALADAPLA"  
gene 1598..2578  
/gene="VNG1213C"  
1598..2578  
/gene="VNG1213C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1213C"  
/protein\_id="AAG19581.1"  
/db\_xref="GI:10580745"  
/translation="MSAAQOTVHRSAQSGRASDIGNPLPTATFDRLHCHATA  
PTSCAPATERASAPRRETELPMCOPOAVMRVENSFLPAAGVETERLQMCHV  
THMDFDGGDGIQDANTENISFIDAERLADAGNTSPFADALPSGELMRVENVAD  
DVAEPDITGGDQSSSVTVTVSVAHGAGDTTLVNGSDLTARLQALLDAPVAVAG  
KRFDAPLEHFPDLSLDSPHLDLMPCRDLDTGGLKOVERDVGVRABVDVDRGA  
VRLMRVABDNDEALDRLVTVNYQDTQLAIDAVERLRHVDVEEPHV"  
gene 3236..4366  
/gene="VNG1214H"  
3236..4366  
/gene="VNG1214H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1214H"  
/protein\_id="AAG19582.1"  
/db\_xref="GI:10580746"  
/translation="WTSHTDRLAEIEDPPLERLAIDLITRTDRIQGVDPQGRKDG  
KXGGLDGPDSNTIIVHSRREDKQKLTDLGAADHRDYDI FVYVTVNIIITGNQ  
KVPVPAQPFVDEYGEIDIMDGERLSELNHHQDLREYLRILADEDPGKTA  
DERLLRRDNELEPRVADAPTAVLHLPVHPAVSGDTDFRHDLPTSPGKRYGYS  
YENTLDGRVTYARGGDNEPDREGYIVDTGEGWEAVDAMGNEENMTIGGOSPEK  
AYEYGRNALNELDGPFEVGLSVLSYKGYSATYKKGCGFDRNGKVRONDIEAKPH  
TVEEDTSPTEGAKRGFDRVMKARWSDGSPYSIDNEWHEF"  
gene 4565..4638  
/gene="VNG1215G"  
4565..4638  
/gene="VNG1215G"  
/note="synonym: VNG1215G"  
complement(4650..5132)  
/gene="pai1"  
/note="pai1"  
/codon\_start=1  
/transl\_table=11  
/product="sporulation regulator homolog"  
/protein\_id="AAG19583.1"  
/db\_xref="GI:10580747"  
/translation="WTVRPAAGTDIDANAEIWLQAAQRAHGSLLAAENHAQARDL  
IGQYVHADGVAITQAGRTVGVMHTGTGFEYEDTRGVINDLTVAGPFGDGGSA  
LTHAERELRSRGAALAI EALADNAARRLRYERGVA  
PHRVTFEOSTVENDTJHSDP

gene complement(5129..6325)  
/gene="pgk"  
/note="synonym: VNG1216G"  
CDS complement(5129..6325)  
/gene="pgk"  
/note="pgk"  
/codon\_start=1  
/transl\_table=11  
/product="glucose-6-phosphate isomerase"  
/protein\_id="AAG19584.1"  
/db\_xref="GI:10580748"  
/translation="WAKRTDLDLAANPAIGRVNDINSPLTRAGGLADARLRAYDT  
LAELLAAPAAVAVLHQRPGDEPARLERHADRLDALDPAVSYCDATFFSGADAV  
ADLAGEVIVLENTFESEYWAFAPEAAADPTALVDGLAPADAYVNAFAAHRSP  
SLVGPETLPSYAGWEAEIDLSGVADPTPTPTPTVYVGAAPVPSVEVAALSHGL  
ADNVTWGVNENPLAAGVNDIGASTPTDPTHERDYGTEIARAADILAAHNDALHLPV  
DAVERDGRCEIETDLPAGDEAVCDIGSDTVDA YAAVYLADESITYVNVGPGVGED  
LFADTRGVFPDAISEVHSIVGGGDTAAIRFDITGFDHSTGGAAINILTDALP  
AVALR"  
gene complement(6428..7144)  
/gene="VNG1218C"  
6428..7144  
/gene="VNG1218C"  
complement(6428..7144)  
/gene="VNG1218C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1218C"  
/protein\_id="AAG19585.1"  
/db\_xref="GI:10580749"  
/translation="WQRFAGCTPTAAVMAVALGPPRSVHGFFYRGLYRKNDAVLEB  
ALTSARVTSASAPDPAATLREHDATAAVQRAAGDLAGVTAADAVADAGGLD  
SVADVNAADPVSLSVATVADADIMARIGDGCVLVVDADABAIIVASADAIALA  
ATNDRGRSRLPGSDPTPVGASGAGADPESRDGPTTAGANTYSQGCERCQG  
LAALGRNGRLCTECSL"  
gene 7107..7853  
/gene="urk"  
7107..7853  
/gene="urk"  
/note="urk"  
/codon\_start=1  
/transl\_table=11  
/product="uridine kinase"  
/protein\_id="AAG19586.1"  
/db\_xref="GI:10580750"  
/translation="WTAAGVOPARCIQHASPPRAVSVTIBSFAIGAGTGAGTIV  
ARETIDNGESATLPLDNYTEDSDRFEERANNAYDHPSEFELRTHIDALLSG  
QSIERPQYDFERHAKADRVVVEPTDVLEGITALLSDETVMDMLHIYVETDADR  
ILRRIERDVFERGHELEGVMDQYISTVPMHQFLEPTKRHADIIIPGANSVAVNLL  
BEKQAEISEMTAMARDDTAFEDATVNPQ"  
gene complement(7911..8234)  
/gene="VNG1220H"  
7911..8234  
/gene="VNG1220H"  
complement(7911..8234)  
/gene="VNG1220H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1220H"  
/protein\_id="AAG19587.1"  
/db\_xref="GI:10580751"  
/translation="NAEMEDPSGRDWVDPDGEHSGEGARKYGLAVLSKMTVEGDFP  
LDAAEFVAEYGEYPRLNHQTIVSYADIFHHIDDEFTKVGPHKTVGAALQDADF  
LDPDA"  
gene complement(8533..8615)  
/gene="trn20"  
8533..8615  
/gene="trn20"  
complement(8533..8615)  
/gene="trn20"  
/product="tRNA-Ser"  
8926..10119  
/gene="orc8"  
10119..10119  
/gene="orc8"  
complement(10119..10119)  
/gene="orc8"

```
/note="Orc8"  
/codon_start=1  
/transl_table=1  
/product="orc / cell division control protein 6"  
/protein_id="AA019588.1"  
/db_xref="GI:10580752"  
/translation="MKTFRRRVRLFKKQVILKHVPEPEIREREDEIDQYANALQDV  
VDGMEPDNVFVYGTGKTAIRIMDALFEYEDDDGVDVSVVNCCHHPSXQ  
AAIALVNELRGDTSDPLTGLSTSDVINALFDEIARBGTVLILDEIDNLDDDL  
LYOLPRAKTNGNIEDSQVAVVGIENDYTFRNDLSPKVQDTLCEREIKFPYDANEIY  
ILDDRAERALSGLVLTGAVIPQCAALARDGSAKQALDILRESVNAIEDERTIVE  
DVEYAVRVRERGRKDSIKDLTHGQVVLAVTQTAIADDTPVRAKELYEVYADIAA
```

Query Match	1.1%;	Score 21;	DB 1;	Length 10225;
Best local Similarity	100.0%;	Pred. No. 7.9;		
Matches	21;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

QY	1104	GGCCGAACCGCTCCACCGGGA	1124
Db	2536	GGCCGAACCGCTCCACCGGGA	2556

Search completed: March 30, 2004, 14:48:31  
Job time : 7035.56 secs



TITLE	JOURNAL	COMMENT
Direct Submission	Submitted (01-JUL-2003) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany	
FEATURES	This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone MGML. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: <a href="http://www.mpiz-koeln.mpg.de/GABI-Kat/">http://www.mpiz-koeln.mpg.de/GABI-Kat/</a> .	
SOURCE	Location/Qualifiers 1..395 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /strain="Columbia 0" /db_xref="taxon:3702" /clone="GK-547B01-020631" /clone_lib="Arabidopsis thaliana T-DNA insertion lines" /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"	

## ORIGIN

Query Match	52.5%	Score 364.6;	DB 29;	Length 395;
Best Local Similarity	95.2%;	Pred. No. 2.1e-50;		
Matches 376;	Conservative	0;	Mismatches 19;	Indels 0;
				Gaps 0

QY	114	ACTGATGAAATTTTGGAGTCCAGATCCGAAAAACGAGCCGTTTAAAGCTTAAATAG	173
Db	1	ACTGATGAAATTTTGGAGTCCAGAAATCGAAAAACAGAGCCGTTTAAAGCTTAAATAG	60
QY	174	CTTCCGTAATTTGTCCTTCCTCGTCACTTTATTTCTTCCCGGAGTCCGTACTCACTA	233
Db	61	CTTCCGTAATTTGTCCTTCCTCGTCACTTTATTTCTTCCCGGAGTCCGTACTCACTA	120
QY	234	CTCTCACTCTCCGGCGCTTTAAACTTACGTTCTCCGTCTTACTCTGTAAAGTTTCTGC	293
Db	121	CTCTCACTCTCCGGCGCTTTAAACTTACGTTCTCCGTCTTACTCTGTAAAGTTTCTGC	180
QY	294	CTTACAGCCTCCGATCGCTCAACCGCAATGCAATCTGCTCGATTTCTTTCTTCGCG	353
Db	181	CTTAAAGCCTCCGATCGCTCAACCGCAATGCTGCTCGATTTCTTTCTTCGCG	240
QY	354	TGGAAAAATTTGCCCTAATGTTCTCGATTTCCAGAGTTTGTGCTATGGATTACTTTT	413
Db	241	TGGAAAAATTTGCCCTAATGTTCTCGATTTCCAGAGTTTGTGCTATGGATTACTTTT	300
QY	414	CCCTAATATTTTAAAGTTCTTAAAGTAAACGATACCTCGGCTTAACTGTTTTGTCAATTTG	473
Db	301	CCCTAATATTTTAAAGTCTTAAAGTAAACGATACCTCGGCTTAAATGTTTTGTCAATTTG	360
QY	474	TTGTGCTTTCACCGTTTAGTGGCTATCGGAGTAT	508
Db	361	TTGTGCTTTCACCGATTTTGTGCTATCAGAGAAT	395

RESULT 2	303 bp	DNA	linear	GSS 02-JUL-2003
LOCUS	BX547724			
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-547E01-020586, genomic survey sequence.			
ACCESSION	BX547724			
VERSION	BX547724			
KEYWORDS	BX547724.1	GI:32440533		
SOURCE	GSS.			
ORGANISM	Arabidopsis thaliana (thale cress)			
	Arabidopsis thaliana			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			

```

REFERENCE
AUTHORS      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eustroids II; Brassicales; Brassicaceae; Arabidopsis.
TITLE
JOURNAL
REFERENCE
AUTHORS      Sririzhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H.
              and Weisshaar,B.
TITLE        A pipeline for automated high-throughput generation of FSTs
              (flanking sequence tags) from Arabidopsis thaliana T-DNA
              transformed lines
JOURNAL
REFERENCE
AUTHORS      Rosso,M., Sririzhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
              A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
              for flanking sequence tag based reverse genetics
TITLE        Unpublished
              3 (bases 1 to 303)
JOURNAL
REFERENCE
AUTHORS      Rosso,M., Li,Y., Sririzhov,N. and Weisshaar,B.
              Direct Submission
              Submitted (01-01U-2003) Weisshaar B., Max-Planck-Institut fuer
              Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
              This sequence is recovered from the left border of the T-DNA. It
              indicates an insertion within the locus defined by clone W0M1. The
              sequences are generated at the MPI for Plant Breeding Research in
              the context of the GABI-Kat project. GABI-Kat is part of the German
              Plant Genomics program designated 'GABI'. Information on line
              availability can be found at:
              http://www.mpiz-koeln.mpg.de/GABI-Kat/.
COMMENT
SOURCE        Location/Qualifiers
              1..303
              1..303
              "Arabidopsis thaliana"

```

**FEATURES**  
**SOURCE**

```

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-547E01-020586"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

```

**ORIGIN**

Query Match	39.9%	Score 277.4;	DB 29;	Length 303;
Best Local Similarity	94.7%	Pred. No. 5e-36;		
Matches 287; Conservative	0;	Mismatches 16;	Indels 0;	Gaps 0;

Qy	115	CTGAATGAAATTTGGAGTCCAGATCGGAAAACGAACCGTATTACAGCTTAATTAAGC	1
Db	1	CTGAATGATTAATTTGGAGTCCAGATCGGAAAACGAAGCCGTTTAAAGCTTAATTAAGC	60
Qy	175	TTCTCCATTGTCCTCTCTTCGTCGATTATTTCTTCCTCCGAGTCTGTACTACTAC	234
Db	61	TTCTCCATTGTCCTCTCTTCGTCGATTATTTCTTCCTCCGAGTCTGTACTACTAC	120
Qy	235	TCCTACTCTCCGAGCGTTTAACTTAACGTTCTCCGTCGTTACTGTAAAGTTTCTGCC	294
Db	121	TCCTACTCTCCGAGCGTTTAACTTAACGTTCTCCGTCGTTACTGTAAAGTTTCTGCC	180
Qy	295	TTAGACCTCCGATCGCCTCACCGATGCAATCTGAGTCGATTCCTTTTCTTCGCT	354
Db	181	TTAAACCTCCGATCGCCTCACCGATGCAATCTGAGTCGATTCCTTTTCTTCGCT	240
Qy	355	GGAAAATTTGCCCTAATGTTCTCGATTTGAAAGTTTGTGCTATGSGTTACTTTTTC	414
Db	241	GGAAAATTTGCCCTAATGTTCTCGATTTGAAAGTATTCGCTGTGATTAATCTTTTTC	300
Qy	415	CCT 417	
Db	301	CCT 303	



RESULT 3	BH854770	209 bp	DNA	linear	GSS 08-JUL-2002
LOCUS	SALK_088648.19.55.x Arabidopsis thaliana TDNA insertion lines				
DEFINITION	Arabidopsis thaliana genomic clone SALK_088648.19.55.x, genomic survey sequence.				
ACCESSION	BH854770				
VERSION	BH854770.1	GI:21704360			
KEYWORDS	GSS.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi				
AUTHORS	1 (bases 1 to 209) Alonso,J.M., Leisse,T.U., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmermann,J. and Ecker,J.R. A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome Unpublished (2001)				
TITLE	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel.: 858 453 4100 x1752 Fax: 858 558 6379 Email: eckere@salk.edu This is single pass sequence recovered from the left border of TDNA.				
JOURNAL	Class: TDNA tagged.				
COMMENT	Location/Qualifiers				
FEATURES	1..209				
SOURCE	/organism="Arabidopsis thaliana" /mol_type="genomic DNA" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_088648.19.55.x" /name="TcPcr was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"				
ORIGIN					
Query Match	22.7%; Score 157.8; DB 28; Length 209;				
Best Local Similarity	84.7%; Pred. No. 2.6e-16;				
Matches 177; Conservative	0; Mismatches 32; Indels 0; Gaps 0;				
CY	322 GCATTCGTGCTCCATTCTCTTTTCTTCGCTGGAAAAATGGCCCTAATGTTCTGCATT 381				
DB	1 GCATTCTGTCGCCGATTCGCTTTTGTTCAACAGAAAATGGCCCTAATGTTCTGCATT 60				
CY	382 TCGAAGTTTTGNCCTANNGGTTACTTTTTTCCCTAATTTTAATAGTTCTTAGTAACG 441				
DB	61 TCGAAGTTTTGNCCTANNGGTTACTTTTTTCCCTAATTTTAATAGTTCTTAGTAACG 120				
CY	442 ATACCTGGCTTACTGTTTTGTTGATATTTGTTGCTTCACCGTTAGTCGCTGATC 501				
DB	121 ATGCTGCGTCTTAGTGCTCATGAACAATGGCGGGTGCTTCCCGGTTAGCGCGTGAATC 180				
CY	502 GGAGATTGACTGTAAGAATAATCCTTCGT 530				
DB	181 GGAGATTGACTGTAAGAATAATCCTTCGT 209				
RESULT 4	BH847287	144 bp	DNA	linear	GSS 13-JUN-2002
LOCUS	BH847287				
DEFINITION	SALK_050870.18.25.x Arabidopsis thaliana TDNA insertion lines				

ACCESSION	BH847287	Arabidopsis thaliana genomic clone SALK_050870.18..25.x, genomic survey sequence.
VERSION	BH847287.1	GI:21418158
KEYWORDS	GSS.	
SOURCE	Arabidopsis thaliana	
ORGANISM	Arabidopsis thaliana (thale cress)	
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.	
AUTHORS	Alonso,J.M., Leisse,T.J., Batajras,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shim,P., Zimmerman,J. and Becker,J.R.	
TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Joseph R. Becker Salk Institute Genomic Analysis Laboratory (SIGNAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel.: 858 453 4100 x1752 Fax: 858 558 6379 Email: jbecker@salk.edu This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged. Location/Qualifiers	
FEATURES	1..144	/organism="Arabidopsis thaliana"
source	/mol_type="genomic DNA"	
	/strain="Columbia 0"	
	/db_xref="taxon:3702"	
	/clone="SALK_050870.18..25.x"	
	/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/dna_protocols.html"	
ORIGIN		
Query Match	15.7%; Score 108.8; DB 28; Length 144;	
Best Local Similarity	84.7%; Pred. No. 3.6e+08;	
Matches 122; Conservative	0; Mismatches 22; Indels 0; Gaps 0;	
CY	372 GTTCTGATTTCCGAAGGTCTTTTGCGATAGGGTACTTTTTCCCTAATTATTAAGTC 431	
DB	1 GTTCTCGAGTTGCACGCAATTTTGCTAGGGGCTACTTTTTCCCTAATTTTACACCTC 60	
CY	432 TTAGTAGACGATACCTCGCTTACTGTTTTTGTCATTTTGTGTGCTTACACGTTTA 491	
DB	61 TTAGTAGACGATACCTCGCTTACTGTTTTTGCCCTTTTGGTGGCATTCGCCCGCTGA 120	
CY	492 GTGCGTAGCGAGTATTGACTG 515	
DB	121 GTGCGTAGGGGAGTATTGACTG 144	
RESULT 5		
LOCUS	AO964581	
DEFINITION	LERGX20TR LEERG Arabidopsis thaliana genomic clone LEERGX20, genomic survey sequence.	
ACCESSION	AO964581	
VERSION	AO964581.1	GI:6792282
KEYWORDS	GSS.	
SOURCE	Arabidopsis thaliana (thale cress)	
ORGANISM	Arabidopsis thaliana	
	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	

REFERENCE	1 (bases 1 to 23)	pai, G., Barnsread, M., Bowman, C., Uterbach, T.,
AUTHORS	Buell, C.R., Lin, X.,	
	Robinson, T. and Fraser, C.M.	
	and F. Cressy, T. and Fraser, C.M.	

JOURNAL  
COMMENT  
Published (2000)  
Contact: Xiaoying Lin  
The Institute for Genomic Research

Email: [at@ctigr.org](mailto:at@ctigr.org)  
For additional information, see <http://www.tigr.org/ctdb/at/at.html>  
Seq primer: TR  
Class: shotgun.  
Location/Qualifiers

```

1..232
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="LANDSBERG_EREECTA"
/db_xref="taxon:3702"
/clone="1BERGX20"
/clone_lib="1BERG"
/note="Organ: Leaf; Vector: pUC19K; Total genomic DNA was sheared to 0.4-0.7 kbp before ligation."

```

ORIGIN

Query Match	14.9%	Score 103.4	UB 20	Length 227
Best Local Similarity	94.7%	Pred. No. 2e-07		
Matches 107; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0

QY 583 NGATTGTTTTCAGCCCTATGCGTTGTGGCTGTGTATTAACCTCAACGTTCAATGCTGAT 64

Db 16 TGCCTGGAAGTTAGCCCTATGCGTTGTGGCTGTGTATTAACCTCAACGTTCAATGCTGAT 75

QY 643 TTTGAGATTTTCGATGCACTGAGGGTTTCTTTGGGCGCTATAGATTTGAAA 695

Db 76 TTTTAGATTTTGGTAGTCACTGAGGGTTTCTTTGGGCGCTATAGATTTGAAA 128

RESULT 6					
CF872119					
LOCUS	CF872119	706 bp	mRNA	linear	EST 31-OCT-2003
DEFINITION	tr1c029xp16.b1 T. reesei mycelial culture.				Version 6 October 2003
	Hyocortra_tecorina cDNA clone tr1c029xp16.		mRNA sequence.		

KEYWORDS	REFERENCE
SIST.	
Source	Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM	Hypocrea jecorina
	Bukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes
	Hypocreales; Hypocreales; Hypocreataceae; Hypocrea
	1 (bases 1 to 706)
AUTHORS	Diemer, S.E., Danneberg, L., Dunn-Coleman, N., Houfek, T.D., Collins, P., Tennissen, P.J.M., Ward, M., and

**JOURNAL COMMENT**  
Unpublished (2003)  
Contact: Ralph A. Dean  
Bureau of Economics Laboratory

```

FEATURES
  source
    location/Qualifiers
      1. 706
        /organism="Hypocrea jecorina"
        /mol_type="mRNA"
        /strain="Qm6a"

```

```

/db_xref="taxon:51453"
/clone="tric029p1c"
/dev_stage="mycelia"
/clone_id="1.1.reesei mycelial culture, Version 6 October 2003"
/notes="vector: PREPty, Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

```

ORIGIN

Query Match	13.2%;	Score 92;	DB 14;	Length 700;
Best Local Similarity	45.9%;	Pred. NO. 6.9e-06;		
Matches 308; Conservative	0;	Mismatches 363;	Indels 0;	Gaps 0

QY	5	ABCAAACAAAATTAAAGCAAGTCATGTTGTAACAAATTAATTAATGTGGAAACAATAA	83
Db	4	AA	63
QY	65	GTAAAGCAAAAAGAAAAAAGAGTACAAAATGAAAAACAATCAACTGAATGAA	124
Db	64	AA	123
QY	125	ATTGGAGTCCAGATCCGAAAAAAGAGCGCTTTAGAGCTTAATAGCTCTCATTT	184
Db	124	AAAGGGGGCGCGGGTTTCATTTTTTTTTTTTTTTTTTTTTCTTTTTCTTTTCTCTCTT	183
QY	185	GTCTCTTCTGTGTCAGTTATTTTTCTTCTCCGAGATCTGACACACTGACCTC	244
Db	184	TTCTTTTTCTTTTTTTTTT	243
QY	245	CGAGCGTTAAACTTAACGTTCGCGTGTACTCTGTAAAGTTTCTGCTTAAGCCTC	304
Db	244	TTTTCTTTCTTTTTTTTTTTTTTTTTTTTTCTTTTTCTTTCTTTCTCTTTTTTTTTTC	303
QY	305	CGATGCGCTCACGGAATGCAATCTGCTGCATTTTCTTTTTCTGCGTGGAAAAATG	364
Db	304	TTTC	363
QY	365	CCCTAATGTTTCGATTTGGAAGTTTTGTGCTAATGGTAACTTTTTCCCTAATATT	424
Db	364	CTTTTTTCTTTTTTTTTTTTTTTTTCTTCTCTCTTTTTTTTTTTTTTTTTTTTTTTTTT	423
QY	425	ATAGTCTTAAGTAACGATACCTGCGCTTACTGTTTGTTCATTTGTGTGAGCTTCA	484
Db	424	TTTTTTTTTTTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTTTTTTTTT	483
QY	485	CCGTTAGTCGCTGATCCGAGTAATGACTGTGAAAAATCCTTGTTTTGGTTTTGT	544
Db	484	TTCTTCTCTTCTCTCTTTTTTTT	543
QY	545	TTCAATATAATCGAGTATCTACCTTTGTGCTTGAATGTTGTTTTAGACCTATGC	604
Db	544	TTTTTTTTTCTTTNTTTTTTTTTTTTTCTTCTCTCTTTTTCTTTTTCTTTTTTTTTT	603
QY	605	GTGTGGCTGTATACTCAAGTCAATGTGGAATTTGAGATTTGTATGATGACTG	664
Db	604	TTTTTTTTCTTTTTCTTCTCTTTTTTTTTTTTTCTTTTTTTTTCTTTCTTTCTTTT	663
QY	665	TGGGTTCTTT 675	
Db	664	CTTTTTTTTT 674	

RESULT 7						
CNS0073W	CNS0073W	922 bp	DNA linear	GSS 03-JUN-1999		
LOCUS	Drosophila melanogaster genome survey sequence TERTJ end of BAC #					
DEFINITION	BACRA4D09 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
ACCESSION	AL066784					
VERSION	AL066784.1 GI:4945247					
KEYWORDS	GSS.					
SOURCE	Drosophila melanogaster (fruit fly)					



Matches	54; Conservative	185; Mismatches	129; Indels	0; Gaps	0;
QY	324	ATTCGTGCTCGATTCTCTTTTCTTTCGCTGAAAAAATGCCCCAATGTTCTCGATTTC	383		
Db	330	MMMTMTMTT	389		
QY	384	GAAGTTTTGTGCTAAGGGTACTTTTTTCCCAATATTATAGTCTTAAGTAACGAT	443		
Db	390	TTTTTTTTTTTTTKTKTKTTTTTTTTTTTTTTTTTTTTTTTTTKTKTKTKTKTKTKTK	449		
QY	444	AACGCGCTTACTGTTTTTGTTCATTGTGTGCTTTCACCGTTTAGTGCAGATCG	503		
Db	450	KTTTKTKTKKKKKTKTKTKTKTKGKKKKKKTKKKKKKKKKKKKKKKKKKKKKKKKK	509		
QY	504	AGTATTGACGTGA AAAATCCCTGTTTTGTGTTTGGTTTGCATATTAATCGATTGA	563		
Db	510	KK	569		
QY	564	TCTACCTTTTNGCTTGATGTTGTTTTTGAACCTAAGCTTGCTGTGATTAAC	623		
Db	570	KK	629		
QY	624	TTTCACTGATGCTGGAATTTGAGATTTTGGTAGTACGTGGGTTCTTTGGTGCTA	683		
Db	630	KK	689		
QY	684	TAGTTGT	691		
Db	690	KKKKKKKK	697		

RESULT 10	613 bp	DNA	linear	GSS 28-JAN-2000
AO964580				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

AO964580  
 LERGX20TF LERGX Arabidopsis thaliana genomic clone LERGX20, genomic  
 survey sequence.  
 AO964580  
 AO964580.1 GI:6792281  
 GSS.  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Burkholderia, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots;  
 rosids, eurosids II; Brassicales, Brassicaceae, Arabidopsis.  
 1 (bases 1 to 613)  
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utterbach, T.,  
 Feldblyum, T., Liang, F., Creasy, T., and Fraser, C.M.  
 Genomic survey sequencing of landsberg erecta ecotype of  
 Arabidopsis thaliana and identification of sequence-based  
 polymorphisms  
 Unpublished (2000)  
 Contact: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atc@igf.org  
 For additional information, see <http://www.tigr.org/cdb/ac/at.html>  
 Seq primer: TP  
 Class: shotgun.

```

FEATURES
    source
        location/Qualifiers
            1..613
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /strain="LANDSBERG ERECTA"
            /db_xref="taxon:3702"
            /clone="LERGX20"
            /clone_1lb="LERG"
            /note="Organ: Leaf; Vector: pUC19/k; Total genomic DNA was
            sheared to 0.4-0.7 kbp before ligation."

```

Query Match	12.2%;	Score 84.8;	DB 28;	Length 6137
-------------	--------	-------------	--------	-------------

Best Local Similarity 97.7%; Pred. No. 0.00012;  
Matches 86; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 608 GTTGGCTTTAACTTCAACGCTGATGATTTAGATTTGTAGTACGCTGG 667  
Db 613 GTTGGAGTGTAACTTCAACGCTGATGATTTGTAGTATTTGTAGTACGCTGG 554

QY 668 GTTCTTTGTGCTATAGTGTGTAATAA 695  
Db 553 GTTCTTTGTGCTATAGTGTGTAATAA 526

RESULT 11  
LOCUS AU236368 647 bp mRNA linear EST 01-APR-2002  
DEFINITION AU236368 RAF14 Arabidopsis thaliana cDNA clone RAF14-93-K05 5',  
mRNA sequence.  
ACCESSION AU236368.1 GI:19875537  
VERSION AU236368  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 647)  
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,  
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,  
Arakawa, T., Shibata, K., Shinagawa, A., Mutamatsu, M., Hayashizaki, Y.,  
Shinozaki, K.  
Large scale analysis of Arabidopsis full-length cDNA (2002b)  
Unpublished (2002)  
Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@rtc.riken.go.jp  
An Arabidopsis full-length cDNA library was constructed essentially  
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
and XhoI was ligated to modified Lambda FL-C1 vector (Carninci et  
al., submitted for publication) digested with BamHI and SalI. This  
clone is in a modified pBluescript vector. Please visit our web  
site: (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further  
details.

FEATURES  
source location/Qualifiers  
1..647  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAF14-93-K05"  
/issue\_type="root"  
/lab\_host="DH10B"  
/clone\_lib="RAF14"  
/note="Site\_1: BamHI; Site\_2: SalI"

ORIGIN  
Query Match 12.1%; Score 84; DB 9; Length 647;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 AATTATTTTCTCCCGGAGTCCGACTCACTCACTCTCCCGGCGTTTAACT 258  
Db 1 AATTATTTTCTCCCGGAGTCCGACTCACTCACTCTCCCGGCGTTTAACT 60

QY 258 TACGTTCTCCGCTTACTCTGT 282  
Db 61 TACGTTCTCCGCTTACTCTGT 84

RESULT 12  
B2558518

LOCUS B2558518 1872 bp DNA linear GSS 17-DEC-2002  
DEFINITION pa98401.209.x1 pac62-164 Pseudomonas aeruginosa genomic clone  
pa98401.209, genomic survey sequence.  
ACCESSION B2558518  
VERSION B2558518.1 GI:27173068  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 1872)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Haefliger, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: Shotgun.

FEATURES  
source location/Qualifiers  
1..1872  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pa98401.209"  
/clone\_lib="pac62-164"  
/note="Clinical isolate 2-164 Whole genomic shotgun  
library."

ORIGIN  
Query Match 11.9%; Score 82.8; DB 28; Length 1872;  
Best Local Similarity 46.2%; Pred. No. 0.00012;  
Matches 304; Conservative 0; Mismatches 348; Indels 6; Gaps 1;

QY 2 ACAAACAAACAAATTTAAGCAAGTCATGTCGTAATTAATTAATGCGAACAAT 61  
Db 527 AAAAAAAGGCGGAAAAAGGCGGAAAAAGGCGGAAAAAGGCGGAAAAAGGCGG 586

QY 62 TAAAGTTAAGCGAAAAAGGAAAAAGGTAAGAAAAAGGAAAAAGGAAAAAGGAAAG 121  
Db 587 AAAAAAAGGCGGAAAAAGGAAAAAGGTAAGAAAAAGGAAAAAGGAAAAAGGAAAG 646

QY 122 AAAAAATTTGAGTCAGAAATCGAAAAACGAGCGCTTTTAAAGCTTAATTAAGCTTCTCA 181  
Db 647 AAAAAAGGCGGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGG 706

QY 182 TTTGCTCTCTTCTGCTCAAGTTATTTCTTCTCGGAGTCTGACCTCACTCTCACT 241  
Db 707 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 766

QY 242 CTCGGCGCTTAACTTAAGTCTCCGCTTACCTGTAAGTTTCTGCTTAAGAC 301  
Db 767 TGTGTGTT 826

QY 302 CTCGATGCGCTCACCGACATGATCTGATCTGATCTCTCTCTCTCTCTCTCTCTCTCTCT 361  
Db 827 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 886

QY 362 TTGCCCCAATGTTCTGATTTGAAAGTTTGTGCTAAGGTTAATTTTCCCTAAT 421  
Db 887 TTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 946

QY 422 TTTAATGTTCTAAGTAAGATACCTGCTTACTGTTTGTGTCATTTGTGTCGCT 481  
Db 947 TTTTATTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1006

482 TCACGTTTAAAGCTGATCGAGATTTGAATGAAATCAATCTTCTGTTTGTGTTT 541

[illegible]

Salk Institute Genomic  
Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: eckerk@salk.edu  
This is single pass sequence recovered from the left border of  
TDNA.  
Class: TDNA tagged.  
Location/Qualifiers  
1..61  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"

/db xref="taxon:3702"  
/clone="SALK\_061198.45.85.x"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tDNA\\_protocols.html](http://signal.salk.edu/tDNA_protocols.html)"

## ORIGIN

Query Match 11.7%; Score 81; DB 28; Length 81;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ACTTCCTCATTTGCTCTCTCCGACATTTATTTCTCCCGAGATCCGACTAC 231  
DB 81 ACTTCCTCATTTGCTCTCTCCGACATTTATTTCTCCCGAGATCCGACTAC 22  
QY 232 TACTCTCACTCTCCGCGCTT 252  
DB 21 TACTCTCACTCTCCGCGCTT 1

RESULT 15  
AL514657/c 1201 bp mRNA linear EST 08-MAY-2003  
LOCUS AL514657 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone  
DEFINITION CLOBB002ZE09 3-PRIME, mRNA sequence.  
ACCESSION AL514657  
VERSION AL514657.2 GI:30464542  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On Feb 13, 2001 this sequence version replaced gi:12778151.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr); Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6668.f For  
more information about this cluster, see  
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CLOBB002ZE09FP1&cluster=6668.f)  
Peng Liang Email : [fliang@life.techn.com](mailto:fliang@life.techn.com) URL :  
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>  
Paradey Avenue Genoscope sequence ID : CLOBB002ZE09FP1.  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CLOBB002ZE09"  
/issue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."

## FEATURES

source  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CLOBB002ZE09"  
/issue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."

## ORIGIN

Query Match 11.6%; Score 80.4; DB 9; Length 1201;  
Best Local Similarity 40.0%; Pred. No. 0.00039;  
Matches 220; Conservative 50; Mismatches 280; Indels 0; Gaps 0;

QY 142 GGAAAAACGAGCGCTTTAGAGCTTAATAGCTTCCTCATTTGCTCTTCTTCGTCAGT 201

DB 842 GGAAAAACGAGCTTCGCGCMCAAAAAMKGRARAHCTTTTGTGTTTCRGVGGGGR 783  
QY 202 TTAATTTCTCTCCGAGTCTGACTGACTGCTGACTCTCCGCGCTTAACCTTAC 261  
DB 782 GBAATTTTCTTGTGCRATTTTCTTTTTCCTCCCAAGAGGAGKTTGKGRAGAR 723  
QY 262 GTTTCCTCGCTTACTCTGTAAGTTTCTGCTTAGAGCTCCGATCGCTCACCGCAT 321  
DB 722 GGVCCGCCCHHNCMGSCCCCMKTTTGTGRCCCTGAGCCGCCCTGSCCCCTTWT 663  
QY 322 GCATTCTGTGCTGATTTCTCTTTTCTGCTGAGAAAATGCCCCAATGTTCTGATT 381  
DB 662 TTTTCTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 603  
QY 382 TCGAAGTTTGTGCTAGGTTACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 441  
DB 602 TGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 543  
QY 442 ATACCTGCTTCTTCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 501  
DB 542 TTTTCTTCTTCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 483  
QY 502 GAGATTTTGAAGTGAAGAAATCTTCTTTTGTGTTTGTGTTTGTGTTTGTGTT 561  
DB 482 TTTTCTTCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 423  
QY 562 GATCTACCTTTGTGCTTGTGATGTTGTTTGTGAGCCATGCGTTGTGTTTATA 621  
DB 422 TKTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 363  
QY 622 ACTTCAGTTCATGTTGATTTTGAATTTGTGATGATGATGATGATGATGATGAT 681  
DB 362 GCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 303  
QY 682 TATAGTTGT 691  
DB 302 TTTTCTTTT 293

Search completed: March 30, 2004, 03:07:25  
Job time : 1617.11 secs



**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 16:31:11 ; Search time 52.2914 Seconds  
(without alignments)  
7375.802 Million cell updates/sec

Title: US-10-030-829-1\_COPY\_1\_695  
Perfect score: 695  
Sequence: 1 gacaacaacaataa.....gtggctatagtgtaaa 695

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.6	12.3	7218	1	US-08-232-463-14 Sequence 14, Appl
2	52.6	7.6	7286	3	US-09-331-581-3 Sequence 3, Appl
3	52.6	7.6	7938	3	US-09-331-581-14 Sequence 14, Appl
4	50.6	7.3	396	4	US-09-640-173-53 Sequence 53, Appl
5	50.6	7.3	396	4	US-09-713-550-53 Sequence 53, Appl
6	50.6	7.3	10619	4	US-10-204-708-4 Sequence 4, Appl
7	50.2	7.2	5562	4	US-10-204-708-63 Sequence 63, Appl
8	49.4	7.1	19124	2	US-08-487-826B-13 Sequence 13, Appl
9	47.8	6.9	2447	2	US-09-014-969-14 Sequence 14, Appl
10	47.4	6.8	5666	4	US-10-204-708-29 Sequence 29, Appl
11	47.4	6.8	8607	4	US-10-204-708-72 Sequence 72, Appl
12	47	6.8	8961	4	US-10-204-708-79 Sequence 79, Appl
13	46.8	6.7	19233	4	US-10-204-708-45 Sequence 45, Appl
14	46	6.6	7218	1	US-08-232-463-14 Sequence 14, Appl
15	45.6	6.6	289	3	US-09-007-005-17 Sequence 17, Appl
16	45.6	6.6	289	3	US-09-244-796-17 Sequence 17, Appl
17	45.4	6.5	11049	4	US-10-204-708-23 Sequence 23, Appl
18	44.2	6.4	1493	1	US-08-340-820-24 Sequence 24, Appl
19	44.2	6.4	1493	1	US-08-593-535-24 Sequence 24, Appl
20	44	6.3	240	1	US-08-628-417-6 Sequence 6, Appl
21	44	6.3	2394	4	US-09-800-729-33 Sequence 33, Appl
22	43.6	6.3	359	4	US-09-621-976-16008 Sequence 16008, A
23	43.6	6.3	2915	4	US-09-336-115C-5 Sequence 5, Appl
24	43.6	6.3	5501	4	US-10-204-708-38 Sequence 38, Appl
25	43.4	6.2	140	1	US-08-628-417-5 Sequence 5, Appl
26	43.4	6.2	359	4	US-09-621-976-16019 Sequence 16019, A
27	43.4	6.2	362	4	US-09-621-976-16010 Sequence 16010, A

28	43.4	6.2	365	4	US-09-621-976-16042 Sequence 16042, A
29	43.4	6.2	5455	4	US-10-204-708-33 Sequence 33, Appl
30	43.4	6.2	11015	4	US-10-204-708-55 Sequence 55, Appl
31	43	6.2	578	4	US-09-602-877A-95 Sequence 95, Appl
32	43	6.2	8607	4	US-10-204-708-72 Sequence 72, Appl
33	43	6.2	8961	4	US-10-204-708-80 Sequence 80, Appl
34	43	6.2	19124	2	US-08-487-826B-13 Sequence 13, Appl
35	42.8	6.2	1114	4	US-09-152-060-41 Sequence 41, Appl
36	42.8	6.2	11050	4	US-10-204-708-85 Sequence 85, Appl
37	42.8	6.2	11331	4	US-10-204-708-27 Sequence 27, Appl
38	42.6	6.1	240	1	US-08-628-417-6 Sequence 6, Appl
39	42.6	6.1	371	4	US-09-621-976-16048 Sequence 16048, A
40	42.6	6.1	10144	4	US-10-204-708-94 Sequence 94, Appl
41	42.4	6.1	270	2	US-08-520-678A-30 Sequence 30, Appl
42	42.4	6.1	270	3	US-08-897-126-30 Sequence 30, Appl
43	42.2	6.1	5844	4	US-10-204-708-89 Sequence 89, Appl
44	42.2	6.1	8866	4	US-10-204-708-20 Sequence 20, Appl
45	42	6.0	357	4	US-09-621-976-16058 Sequence 16058, A

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO.: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpc-F18  
US-08-232-463-14  
Query Match 12.3%, Score 85.6, DB 1, Length 7218;

Best Local Similarity 6.4%; Pred. No. 1.4e-12;  
Matches 25; Conservative 232; Mismatches 131; Indels 0; Gaps 0;

Oy	174	CTTCCTCATTTGTCCTTCTTGCGACAGTTAATTCTTCCTCCGGAGCAGCACTA	233
		::::: :::: :::: ::::: ::::: ::::: ::::: ::::: :::::	
Db	1094	YYY	1153
Oy	234	CTCTCACCTCCGGGCTTTAACTTAAGTCTCCGCTTACTCTGAAGTTTTGC	293
		::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::	
Db	1154	YYY	1213
Oy	294	CTTAGAGCTCCGATCGCCTCAACGCATGATCTGTGCTGATTTCTTTCTGCG	353
		::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::	
Db	1214	YYY	1273
Oy	354	TGGAAAAATGCCCTTAAGTTCTGATTTCCGAAGTTTTGNCATGAGTACTTTT	413
		: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::	
Db	1274	YYY	1333
Oy	414	CCCTATTTTATAAGTCTTAAGTAACGATACCTCGCTACTGTTTGTCAATTTG	473
		::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::	
Db	1334	YYY	1393
Oy	474	TTTGTCCTTACCGTTTACGCTGATCGGAGATTTTGACTGAAAAATCCTTGCTT	533
		::: ::::: ::::: ::::: ::::: ::::: ::::: :::::	
Db	1394	YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATCTCTATCT	1453
Oy	534	TTGGTTTTTGTCTATATAATCGAATT	561
Db	1454	CTTTAACTACTGCAATGATAGTAATTT	1481

```

, RESULT 2
, US-09-331-581-3
, Sequence 3, Application US/09331581
, Patent No. 6130070
, GENERAL INFORMATION:
, APPLICANT: TOHDA, Hideki
, APPLICANT: HAMA, Yuko
, APPLICANT: KIMAGAI, Hiromichi
, TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
, TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
, TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
, FILE REFERENCE: 0059-1142-OPT
, CURRENT APPLICATION NUMBER: US/09/331,581
, CURRENT FILING DATE: 1999-06-30
, EARLIER APPLICATION NUMBER: PCT/JP98/04929
, EARLIER FILING DATE: 1998-10-30
, EARLIER APPLICATION NUMBER: JP 9-314608
, EARLIER FILING DATE: 1997-10-31
, NUMBER OF SEQ ID NOS: 24
, SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 3
, LENGTH: 7286
, TYPE: DNA
, ORGANISM: Artificial Sequence
, FEATURES:
, OTHER INFORMATION: Description of Artificial Sequence:DNA
, US-09-331-581-3

```

Query	Match	Similarity	7.6%	Score	52.6	DB	3	Length	7286
Best	Local	Similarity	56.7%	Pred	No.	0.00044			
Matches	97	Conservative	0	Mismatches	74	Indels	0	Gaps	0
Qy	9	AAACAAATTTAAGCAAGTCATGTCGTAGCAATAATTAATATGTGGACAAATTAAGTTA	68						
Db	4559	AACAAATCATTGAGAAAGATGCTTAATGTAGCTGAATAATAATGACGTACAAACCAAAAA	4618						
Qy	69	AGCGAAAAAGAAAAAAAGGTACAAAAATGAAAAACAAATCAAACTGGAATGAAAAATT	128						
Db	4619	AAG	4678						
Qy	129	GGAGTCCAGAAATCGAAAAACGAGCGCGTTTAAAGACTTAATTAAGTTCCT	179						

Db 4679 TACCTTCTGAGCGGAAGAACAGCCGATCCAGACATGATTAAGATACAT 4729

```

US-09-331-581-14
RESULT 3
US-09-331-581-14
Sequence 14, Application US/09331581
Patent No. 6130070
GENERAL INFORMATION:
APPLICANT: TOMODA, Hideki
APPLICANT: HAMA, Yuko
APPLICANT: KIMAGAI, Hiromichi
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
FILE REFERENCE: 0059-1142-0PCT
CURRENT APPLICATION NUMBER: US/09/331,581
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/JP98/04929
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: JP 9-314608
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 7938
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14

```

Query Match	7.6%	Score 52.6	DB 3.6	length 7938
Best Local Similarity	56.7%	Pred. No. 0.00045		
Matches	97	Conservative	0	Mismatches 74
			Indels	0
			Gaps	0
QY	9	AACAAAAATTAGCAGTCATGTTCTGTGACAAATTAATTAAATGAGGACAACTTAAGTTA	68	
DB	5211	AACATATCATGAGAAAGATGTCCTATGTGTCGAAAAATTAATGACGTACACAGACAAAA	5270	
QY	69	ACGAAAAAGGAAAAAAAGGTACAAAATAATGAAAAACAAATCAAACCTGAAATGAATAATTT	128	
DB	5271	AAG	5330	
QY	129	GGAGTCCGAATCGGAAAAACGAGCGCGTTTAAAGCTTAAATTAAGCTTCTT	179	
DB	5331	TACCTTCGAGCGGAAAGACACAGCCCGATTCACATGATGATATCAT	5381	

```

RESULT 4
US-09-640-173-53
; Sequence 53, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiequnhu
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640.173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-53

```

Query Match 7.3%; Score 50.6; DB 4; Length 396;  
Best Local Similarity 49.8%; Pred. No. 0.00049;  
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 448  
DB 18 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 77  
QY 449 CGTCTACTGTTTGTGTCATTTGTCCTTACCGTTAGTCGATGAGAT 508  
DB 78 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 137  
QY 509 TTGACTGGAATAACCTTCGTTTGTGTTTGTTCATATAATCGATGATAC 568  
DB 138 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 197  
QY 569 CTTTGTGCTTGATGTTGTTTGTGACCCATGCGTTGGCTGTATTAATCA 627  
DB 198 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 256

## RESULT 5

US-09-713-550-53  
Sequence 53, Application US/09713550  
Patent No. 6617109

## GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun  
APPLICANT: Stolk, John A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT OF INFECTIONS: THERAPY AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.484C4  
CURRENT APPLICATION NUMBER: US/09/713,550  
CURRENT FILING DATE: 2000-11-14  
NUMBER OF SEQ ID NOS: 205  
SOFTWARE: FastSeq For Windows Version 3.0  
SEQ ID NO 53  
LENGTH: 396  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(396)  
OTHER INFORMATION: n = A,T,C or G  
US-09-713-550-53

Query Match 7.3%; Score 50.6; DB 4; Length 396;  
Best Local Similarity 49.8%; Pred. No. 0.00049;  
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 448  
DB 18 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 77  
QY 449 CGTCTACTGTTTGTGTCATTTGTCCTTACCGTTAGTCGATGAGAT 508  
DB 78 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 137  
QY 509 TTGACTGGAATAACCTTCGTTTGTGTTTGTTCATATAATCGATGATAC 568  
DB 138 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 197  
QY 569 CTTTGTGCTTGATGTTGTTTGTGACCCATGCGTTGGCTGTATTAATCA 627  
DB 198 TTTTGTGATAGGCTTATTTTCCCTATTTATATGTTCTTAGGTAACGATACCG 256

## RESULT 6

US-10-204-708-4  
Sequence 4, Application US/10204708  
Patent No. 6677731

GENERAL INFORMATION:  
APPLICANT: Olek, Alexander  
APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
TITLE OF INVENTION: By Assessing DNA Methylation  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 4  
LENGTH: 10619  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-204-708-4

Query Match 7.3%; Score 50.6; DB 4; Length 10619;  
Best Local Similarity 49.8%; Pred. No. 0.0017;  
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 404 TTAATTTTCCCTATTTATATGTTCTTAGGTAACGATACCGTCTTACTGTTT 463  
DB 9137 TTAATTTTCCCTATTTATATGTTCTTAGGTAACGATACCGTCTTACTGTTT 9196  
QY 464 GTTCAATTTTGTGCTTACCGTTAGTGTGCTATGAGATATGAGTGAAT 523  
DB 9197 TTTGCTTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 9256  
QY 524 CTTGCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 583  
DB 9257 AGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 9316  
QY 584 GTTGTGTTTGTGAGCTTATGCTTGTGCTTGTATTAATCAATGATGATGAT 643  
DB 9317 GTTGTGTTTGTGAGCTTATGCTTGTGCTTGTATTAATCAATGATGATGATGAT 9376  
QY 644 TTGAGATTTTGTGATG 660  
DB 9377 TGTATTTATGTTAGTG 9393

## RESULT 7

US-10-204-708-63  
Sequence 63, Application US/10204708  
Patent No. 6677731

## GENERAL INFORMATION:

APPLICANT: Olek, Alexander  
APPLICANT: PIEPENBROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 63  
LENGTH: 5562  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-63

Query Match 7.2%; Score 50.2; DB 4; Length 5562;  
Best Local Similarity 48.1%; Pred. No. 0.0017; Indels 0; Gaps 0;  
Matches 142; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 390 TTTTGGCATGGTACTTTTCCATATTTATAGTTCTTAGAGTAACTGTC 449  
DB 1394 TATTTATTTATTTTCTTTTTCGTTAGAAAGATTTTTCGATTAATTTAT 1453  
QY 450 GTCTTACTGTTTGTTCATTTTGTGCTTTCACCGTTAGTGGCATGAGTAT 509  
DB 1454 ATTTTATTTTATTTAGTTAGTTTATTTTATTTTATTTATTTATTTATTT 1513  
QY 510 TGACTGTGAAATCCTTGGTTTGTGTTTTCATATTAATCGATGATCTACC 569  
DB 1514 TGTACTTATGATTTTATTTTATTTATTTTATTTTATTTTATTTTATTTAT 1573  
QY 570 TTTTGTGCTTGTGATTTGTTTGTGAGCTTAAGCGTTGGCTTTATTAATCTCAG 629  
DB 1574 TTTTGTGTTTGTAGAAAGTTTATTTATTTTATTTTATTTTATTTTATTTTAT 1633  
QY 630 TTCATGTGTGATTTTGAGATTTTGTAGTGAAGTGGTTCTTGTGGCTAT 684  
DB 1634 GATAGTGTATTTTATTTAGAAATTTGTGATTAATGATTTTATTTATTTAT 1688

RESULT 8  
US-08-487-826B-13/c  
Sequence 13, Application US/08487826B  
Patent No. 5993827

GENERAL INFORMATION:  
APPLICANT: Sim. Kim L.

APPLICANT: Chelms, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Welles, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbie Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US

ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487/826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelien, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NH121.001CPI  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:

LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 7.1%; Score 49.4; DB 2; Length 19124;  
Best Local Similarity 47.9%; Pred. No. 0.0042; Indels 1; Gaps 1;  
Matches 172; Conservative 0; Mismatches 186; Indels 1; Gaps 1;

QY 318 GCATGATCTGTCGCGAATTTCTTTTCTTCGCTGGAATAATGCCCTAATGTC 377  
DB 15976 GCATGATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 15917  
QY 378 GATTTGAGATTTTGTGCTATGGTTACTTTTCCATATTTATAGTCTAGCT 437  
DB 15916 TTGATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 15857  
QY 438 AACGATACCTGCTTACTGTTTGTGATTTGTTGCTTTCACCGTTAGTGGCT 497  
DB 15856 TTATTTATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTT 15797  
QY 498 GATGAGATTTGACTGTGAAATCCTTGTGTTTGTGTTTGTTCATATTAATCG 557  
DB 15796 TATGATATATTTTATTTTATTTTATTTTATTTTATTTTATTTATTTAT 15737  
QY 558 G-ATTGATCACTTTTGTGCTTGTGATTTGTTTGTGAGCCATGCGTTGGCTG 616  
DB 15736 TTATTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 15677  
QY 617 TTATTAATCACTGCTGATGATTTGATTTGATTTGATTTGATTTGATTTGAT 675  
DB 15676 AATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 15618

RESULT 9  
US-09-014-969-14/c  
Sequence 14, Application US/09014969  
Patent No. 5963397

GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Treacy, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/014,969  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2447 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-014-969-14

Query Match 6.9%; Score 47.8; DB 2; Length 2447;  
Best Local Similarity 47.2%; Pred. No. 0.0051;  
Matches 142; Conservative 1; Mismatches 158; Indels 0; Gaps 0;

QY 380 TTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGATACCTGC 449  
DB |||||  
QY 450 GCTTACGTTTGTGCTATTTGTGCTTCAACCGTTAGTGCATGCGAGTAT 509  
DB |||||  
QY 2387 TTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGATACCTGC 2388  
DB |||||  
QY 510 TGAAGTGAAGAAATCCTGCTTTTGGTTTGTTCATATATTAACGATGATCTAC 569  
DB |||||  
QY 2327 TTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGATACCTGC 2268  
DB |||||  
QY 570 TTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGATACCTGC 629  
DB |||||  
QY 2267 TTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGATACCTGC 2208  
DB |||||  
QY 630 TTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGATACCTGC 689  
DB |||||  
QY 2207 ATCAATACACGATCCAGATATGTAACCATATATATTAACGATACCTGC 2148  
DB |||||  
QY 690 G 690  
DB 2147 G 2147

RESULT 10  
US-10-204-708-29

Sequence 29, Application US/10204708

Patent No. 6677731

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PIERENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: by Assessing DNA Methylation

FILE REFERENCE: 5013.1012

CURRENT APPLICATION NUMBER: US/10/204,708

PRIOR FILING DATE: 2003-05-06

PRIOR APPLICATION NUMBER: PCT/EP01/03971

PRIOR FILING DATE: 2001-04-06

PRIOR APPLICATION NUMBER: DE 10019058.8

PRIOR FILING DATE: 2000-04-06

PRIOR APPLICATION NUMBER: DE 10019173.8

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: DE 10032529.7

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 98

SEQ ID NO 29

LENGTH: 5666

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-204-708-29

Query Match 6.8%; Score 47.4; DB 4; Length 5666;  
Best Local Similarity 49.2%; Pred. No. 0.0088;  
Matches 153; Conservative 0; Mismatches 156; Indels 2; Gaps 1;

QY 384 GAAGTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGAT 443  
DB |||||  
QY 3358 GTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 3417  
DB |||||  
QY 444 ACCGCGCTTACGCTTTTGTGCTATTTGTGCTTCAACCGTTAGTGCATGCGAGTAT 503  
DB |||||  
QY 3418 TTTTGAAGCTTATTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGAT 3477  
DB |||||  
QY 504 AGTATTTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGAT--T 561  
DB |||||  
QY 3478 TTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGAT--T 3537  
DB |||||  
QY 562 GATCTACCTTTTGTGCTTATGTTTGTGCTTCAACCGTTAGTGCATGCGAGTAT 621  
DB |||||  
QY 3538 GTGTTGCTTCTTTTGTGCTTCAACCGTTAGTGCATGCGAGTAT--T 3597  
DB |||||  
QY 622 ACTTACGCTTATGCTGCTTATGCTGCTTCAACCGTTAGTGCATGCGAGTAT--T 681  
DB |||||  
QY 3558 TGTGTTTGAAGGCTTCTTTTCCCTATATATTAAGTCTTAGAGTAACGAT--T 3657  
DB |||||  
QY 682 TATAGTTGTA 692  
DB |||||  
QY 3658 TATGCTTATA 3668  
DB |||||

RESULT 11  
US-10-204-708-72

Sequence 72, Application US/10204708

Patent No. 6677731

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PIERENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: by Assessing DNA Methylation

FILE REFERENCE: 5013.1012

CURRENT APPLICATION NUMBER: US/10/204,708

PRIOR FILING DATE: 2003-05-06

PRIOR APPLICATION NUMBER: PCT/EP01/03971

PRIOR FILING DATE: 2001-04-06

PRIOR APPLICATION NUMBER: DE 10019058.8

PRIOR FILING DATE: 2000-04-06

PRIOR APPLICATION NUMBER: DE 10019173.8

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: DE 10032529.7

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 98

SEQ ID NO 72

LENGTH: 8607

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-204-708-72

Query Match 6.8%; Score 47.4; DB 4; Length 8607;  
Best Local Similarity 46.7%; Pred. No. 0.01; Indels 0; Gaps 0;  
Matches 150; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 374 TCTGATTCGAAGGCTTCTTTGTGCTAAGGCTTCTTTTCCCTATATATTAAGTCTT 433  
DB |||||  
QY 6966 TATGCTTTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 7025  
DB |||||  
QY 434 AGGTACGATACCTGCGCTTACGTTTGTGCTATTTGTGCTTCAACCGTTAGT 493  
DB |||||  
QY 7026 TTTTGTGCTTCTTTTGTGCTTCAACCGTTAGTGCATGCGAGTAT--T 7085  
DB |||||

QY 494 CGGATGAGAGATTTGACTGAGAAAATCCCTGTTTGGTTTTCATATA 553  
DB 7086 TTTTGTATTTTGTGTTTATATTTTATAGTTGTTTGTGGAAGAA 7145  
QY 554 ATCGATGATGATCCTTTTGTGCTTGAATGTTTGTGACCTATGCTGTGGC 613  
DB 7146 ATAAAGAGGTTTGTGTTGTGTTGTTGTTTGTGTTTGTGTTTGTGTTT 7205  
QY 614 TTGTATACCTTCAGCTTCATGTTGATTTTGAATTTTGTGATGCTGTGCTT 673  
DB 7206 AGAATGATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 7265  
QY 674 TTGGTGCTATAGTTTAA 694  
DB 7266 AGTATGTTTGTGTTGTTAA 7286

RESULT 12  
US-10-204-708-79  
Sequence 79, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
PRIOR FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 79  
LENGTH: 8961  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
FEATURE: chemically treated genomic DNA (Homo sapiens)  
NAME/KEY: unsure  
LOCATION: (3866)  
OTHER INFORMATION: n is a or g or c or t  
US-10-204-708-79

Query Match 6 8%; Score 47; DB 4; Length 8961;  
Best Local Similarity 48.9%; Pred. No. 0.013;  
Matches 155; Conservative 0; Mismatches 160; Indels 2; Gaps 1;  
QY 341 TCTTTCTTCGCTGAAAATGCCCTAATGTTCTCGATTCGAAAGTTTGTGCTAT 400  
DB 6013 TATTGTGTTGATGGGGAAATGTTTATTTTAAATGATGTTGATGTTATTTT 6072  
QY 401 GGGTACTTTTCCCATATTTTATAGTTCTAGTAACATACCTGCTTACTGTT 460  
DB 6073 TAAGTAGATGTTTTTTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 6132  
QY 461 TTGTGATTTTGTGCTTCAACCTTATGCTGATCGGATTTGACTGTA 520  
DB 6133 ATTTTGTGATTTTATGTTTATTTTATTTTATTTTATTTTATTTTATTTT 6190  
QY 521 AATCCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 580  
DB 6191 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 6250

QY 581 GATGTTGTTTGTGACCTATGCTTGTGCTGTTATATCTCAGTTCATGTGCG 640  
DB 6251 ATTTAGTATTTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 6310  
QY 641 ATTTGAGATTTGTGA 657  
DB 6311 TGGGAGATGTTTGA 6327

RESULT 13  
US-10-204-708-45  
Sequence 45, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
PRIOR FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 45  
LENGTH: 19233  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
FEATURE: chemically treated genomic DNA (Homo sapiens)  
NAME/KEY: unsure  
LOCATION: (176, 178, 179, 273, 586, 648, 651, 920, 1014, 1173, 1197, 1228)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1264, 1267, 1559, 1607, 1677, 1703, 1742, 1848, 1850, 2033)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (2168, 2254, 2257, 2280, 2291, 2723, 3018, 3080, 4205, 4209)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (4216, 4225, 4239, 4246, 4254, 4265, 4270, 4274, 4288, 4294...4295)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (4298, 4363, 4610, 4627, 4675, 4693, 4698, 4738, 4781)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (4787, 4825, 4829, 4838, 4855, 5462, 5494, 5496, 5500, 5674)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (5690, 5697, 5705, 5715, 5890, 5910, 5913, 6027, 6080, 6109)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (6244, 6288, 6776, 6873, 6891, 6980, 7028, 7656, 7876, 8030)  
OTHER INFORMATION: n is a or g or c or t  
FEATURE:  
NAME/KEY: unsure



```
/ LOCATION: (8059, 8075, 8632, 8636, 9700, 9717, 9815, 9818, 9867, 9879)
/ OTHER INFORMATION: n is a or g or c or t
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (10005, 10075, 10150, 10159, 10165, 10356, 11123, 11166, 11360)
/ OTHER INFORMATION: n is a or g or c or t
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (11703, 11933, 12050, 12340, 12981, 13564, 13764, 13765, 13787)
/ OTHER INFORMATION: n is a or g or c or t
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (13846, 14168, 16566, 17006, 17217, 17650, 17653, 17656, 17898)
/ OTHER INFORMATION: n is a or g or c or t
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (18130, 18143, 18151, 18188, 18498, 18549, 18936, 18946, 19165)
/ OTHER INFORMATION: n is a or g or c or t
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (19200)
/ OTHER INFORMATION: n is a or g or c or t
/ US-10-204-708-45
```

```
Query Match 6.7%; Score 46.8; DB 4; Length 19233;
Best Local Similarity 48.5%; Pred. No. 0.02;
Matches 129; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
```

```
QY 360 AATGCCCTAATGTTCTGCATTTTCGAAGTTTGTGCTAATGGCTTACTTTTCCCTAT 419
DB 11391 AATTATATGTTTGTGAGATATGATATTTTATTAACGTAAATTTTGTGTT 11450
QY 420 AATTATAGTCTGTAGTACGATACCTCGCTTACTGTTTGTCTATTTGTGTGTC 479
DB 11451 TTTTATTTTATTTTATTTTATTTAGTATTTAGTTTGTGTTTATTTTGTAGGTT 11510
QY 480 TTTTCACTGTTAGTGTGATGAGATATTTGACTGTGAAATACCTGTTTGTGTT 539
DB 11511 AGTTTGTGTTTATTAAGTTTATTAATTAATGAATTAATATTTTGTGTGT 11570
QY 540 TTTGTTTCAATTAATGAGATTTGATCTTACCTTTTGTGCTTGAATGTTTGTGAGCC 599
DB 11571 TTGGTTTTTTTATTTAGTATTTTGTGTTTATTTTGTAGGTTTGTGTTTGTAGGTT 11630
QY 600 TATGGTGTGCTGTTGTTATTAATCTT 625
DB 11631 GGAGTGTGCGATATAGTTATGCTTT 11656
```

```
RESULT 14
US-08-232-463-14/c
/ Sequence 14, Application US/08232463
/ Patent No. 5670367
/ GENERAL INFORMATION:
/ APPLICANT: DORNER, F.
/ APPLICANT: SCHEIFLINGER, F.
/ APPLICANT: FALKNER, F. G.
/ TITLE OF INVENTION: RECOMBINANT FOMLOX VIRUS
/ NUMBER OF SEQUENCES: 52
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Foley & Lardner
/ STREET: 1800 Diagonal Road, Suite 500
/ CITY: Alexandria
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22313-0299
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/232,463
```

```
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/935,313
/ FILING DATE:
/ APPLICATION NUMBER: EP 91 114 300.6
/ FILING DATE: 26-AUG-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BENT, Stephen A.
/ REGISTRATION NUMBER: 29,768
/ REFERENCE/DOCKET NUMBER: 30472/114 IMMU
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 836-9300
/ TELEFAX: (703) 683-4109
/ TELEX: 899149
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 7218 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ CLONE: pTZapc-F18
/ US-08-232-463-14
```

```
Query Match 6.6%; Score 46; DB 1; Length 7218;
Best Local Similarity 5.6%; Pred. No. 0.022;
Matches 10; Conservative 115; Mismatches 55; Indels 0; Gaps 0;
```

```
QY 1 GACAAACAACAAATTAAGCAAGTCACTGTTCTAGCAATTAATTAAGTGGAACAA 60
DB 1231 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1172
QY 61 TTAAGTTAAGCAAAAGGAAAAAGGTCAGAAATGAATGAACAAATCAACTGAAT 120
DB 1171 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112
QY 121 GAAATTTGAGTCAGATCGAATCGAAGGCGGTTTGAAGCTTAATAGCTTCCCTC 180
DB 1111 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1052
```

```
RESULT 15
US-09-007-005-17/c
/ Sequence 17, Application US/09007005B
/ Patent No. 6258558
/ GENERAL INFORMATION:
/ APPLICANT: Szostak, Jack W.
/ APPLICANT: Roberts, Richard W.
/ APPLICANT: Liu, Rife
/ TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
/ FILE REFERENCE: 00786/35003
/ CURRENT APPLICATION NUMBER: US/09/007,005B
/ FILING DATE: 1998-01-14
/ EARLIER APPLICATION NUMBER: 60/035,963
/ EARLIER FILING DATE: 1997-01-27
/ EARLIER APPLICATION NUMBER: 60/064,491
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 17
/ LENGTH: 289
/ TYPE: RNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Translation template
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(289)
/ OTHER INFORMATION: n = A,T,C or G
/ US-09-007-005-17
```

Query Match 6.6%; Score 45.6; DB 3; Length 289;  
Best Local Similarity 10.4%; Pred. No. 0.0085;  
Matches 27; Conservative 105; Mismatches 128; Indels 0; Gaps 0;

```
QY 175 TTCCCTCATTTGCTCTCTCTGAGTTTATTTCTTCTCTCCGAGTCTGACTCAGTAC 234
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 275 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
QY 235 TCTCACTCTCCGCGCCTTAACTTACGTTCTCCGTGTTACTCTGTAAGTTTCTGCC 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 YNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 156
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 295 TTAGACCTCCGATCGCTCACCAGATTCGTGCTCGATTCTCTTTCTTGCT 354
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 YNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 96
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 GGAATAATGCCCTTAATGTTCTCGATTTCGAAGTTTGTGCTAAGGTTACTTTTTC 414
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 YNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 36
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 415 CCTTATTTTATAGTTCTTA 434
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 35 YTVAAVATYATYATYATA 16
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: March 30, 2004, 03:14:52  
Job time : 56.2914 secs

(AVET) AVENTIS CROPS SCIENCE SA.  
(INDC) INCT NAT BECH AGRONOMICITE

Abd346655	Oligonuc1	
Ab133749	Human imm	
Abx52180	Bovine ES	
Abcd21705	Breast can	
Acid923384	Human col	
Aas645677	Tumour su	
Abq452047	Oligonuc1	
Abq52046	Oligonuc1	
Ab195044	Human neu	
Aas646774	Tumour su	
Abq33105	Oligonuc1	
Abq33104	Oligonuc1	
Abq33210	Oligonuc1	
Abq33210	Oligonuc1	
Abq33211	Oligonuc1	
Ab192313	Chemical1	
Abq665981	Human che	
Abn806315	Human che	
Ab133212	Human im	
Abq436604	Oligonuc1	
Abq46605	Oligonuc1	
Abd54216	Preterace	
Abd84152	Human IY	

XX Beclin C, Elmayan T, Vaucheret H;  
XX WPI; 2001-159529/16.  
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
XX resistance in plants and, when inhibited, for increasing transgene  
XX expression.  
XX Claim 1; Page 31-32; 36pp; French.  
XX The present sequence represents the genomic sequence of the Arabidopsis  
XX thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional  
XX inactivation (degradation of RNA) and for resistance to viruses.  
XX Overexpression of SGS3 results in plants with increased resistance to  
XX viruses, while inactivation of SGS3 in transgenic plants (e.g. by  
XX expressing antisense RNA, by mutation or by homologous recombination)  
XX increases the level of the transgene product. This product may e.g.  
XX impart resistance (to herbicide, insects or pathogens), alter contents of  
XX essential fatty acids or proteins, or is pharmaceutically active, e.g. an  
XX immunoglobulin or interferon.

Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;

Query Match 100.0%; Score 695; DB 4; Length 3275;  
Best Local Similarity 100.0%; Pred. No. 3.8e-145;  
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

61 GACAAACAACAAATTTAGCAAGTCATGTCGTAGCAATTAATTAATAGTGGACAA 60  
1 GACAAACAACAAATTTAGCAAGTCATGTCGTAGCAATTAATTAATAGTGGACAA 60  
61 TTAAGTTAAGCAAAAGGAAAAAGGTAACAAATTAATTAATAGTGGACAA 120  
61 TTAAGTTAAGCAAAAGGAAAAAGGTAACAAATTAATTAATAGTGGACAA 120  
121 GAAATTTGAGATCGAAGTGGAAAAAGAGCGCTTTAGAGCTTAATAGTGGACAA 180  
121 GAAATTTGAGATCGAAGTGGAAAAAGAGCGCTTTAGAGCTTAATAGTGGACAA 180  
181 ATTGTCCTCTTCGTCAGTATTATTTCTCTCGAGTCTGACTGACTACTCTGAC 240  
181 ATTGTCCTCTTCGTCAGTATTATTTCTCTCGAGTCTGACTGACTACTCTGAC 240  
241 TCTCCGCGCTTTAAACCTTACGTTCCCTCGTTACTCTGTAAGTTTCTGCTTAAG 300  
241 TCTCCGCGCTTTAAACCTTACGTTCCCTCGTTACTCTGTAAGTTTCTGCTTAAG 300  
301 CCTCGATGCGCTACCGGATGATCTGTCGCTGATTTCTGTTTCTGCTGGAAGA 360  
301 CCTCGATGCGCTACCGGATGATCTGTCGCTGATTTCTGTTTCTGCTGGAAGA 360  
361 ATTGCGCTATGTCGATTCGATTTGAAAGTTTGTGCTATGAGTTACTTTTCCCTATA 420  
361 ATTGCGCTATGTCGATTCGATTTGAAAGTTTGTGCTATGAGTTACTTTTCCCTATA 420  
421 TTTTATAGTTCTTAAAGTAAGTACTGCTGCTTACTGTTTGTGCTATGTTGTGCT 480  
421 TTTTATAGTTCTTAAAGTAAGTACTGCTGCTTACTGTTTGTGCTATGTTGTGCT 480  
481 TTTTATAGTTCTTAAAGTAAGTACTGCTGCTTACTGTTTGTGCTATGTTGTGCT 540  
481 TTTTATAGTTCTTAAAGTAAGTACTGCTGCTTACTGTTTGTGCTATGTTGTGCT 540  
541 TTTTATAGTTCTTAAAGTAAGTACTGCTGCTTACTGTTTGTGCTATGTTGTGCT 600  
541 TTTTATAGTTCTTAAAGTAAGTACTGCTGCTTACTGTTTGTGCTATGTTGTGCT 600  
601 ATGCGTTGTTGCTGTTTAACTTCACTGATGTTGATGTTGATGTTGATGTTGATG 660  
601 ATGCGTTGTTGCTGTTTAACTTCACTGATGTTGATGTTGATGTTGATGTTGATG 660  
661 ACTGTCGTTTCTTGTGCTATAGTGTGTAAGA 695

DB 661 ACTGTCGTTTCTTGTGCTATAGTGTGTAAGA 695

RESULT 2  
ABL33697  
ID ABL33697 standard; DNA; 6668 BP.  
XX ABL33697;  
XX 26-MAR-2002 (first entry)  
XX Human immune system associated gene SEQ ID NO: 1670.  
XX Human immune system associated gene SEQ ID NO: 1670.  
XX Human, immune system disease; cytosine methylation; antiasthmatic;  
XX antiarteriosclerotic; anti-anemic; cytosolic; noctropic;  
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
XX antirheumatic; antiarthritic; antidiabetic; antiparasitic;  
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
XX ds.  
XX Homo sapiens.  
XX OS  
XX WO200200928-A2.  
XX 03-JAN-2002.  
XX 02-JUL-2001; 2001WO-EP007537.  
XX 30-JUN-2000; 2000DE-01032529.  
XX 01-SEP-2000; 2000DE-01043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
XX Nucleic acid comprising fragment of chemically modified gene, useful for  
XX diagnosis and treatment of diseases associated with abnormal cytosine  
XX methylation.  
XX Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.  
XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention

Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 0 U; 1 Other;

Query Match 9.9%; Score 68.8; DB 6; Length 6668;  
Best Local Similarity 45.8%; Pred. No. 1.3e-05;  
Matches 238; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

156 GTTTAAAGCTTAAGCTTCCGATTTGCTCTCTGCGAGTTATTTCTTCTC 215  
2742 GTTTTAAAGCTTAAAGCTTCCGATTTGCTCTCTGCGAGTTATTTCTTCTC 2801  
216 CGAGTCTGACTCACTCACTCTCGGCGCTTAAACTTACGTTCTCGCTGTT 275  
2802 TTTTAAAGCTTAAAGCTTCCGATTTGCTCTCTGCGAGTTATTTCTTCTC 2861  
276 ACTGTAAGTTTCTGCTTAAAGCTTCCGATTTGCTCTCTGCGAGTTATTTCTTCTC 335  
2862 TTTTAAAGCTTAAAGCTTCCGATTTGCTCTCTGCGAGTTATTTCTTCTC 2921



XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1; Page 8100; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

SO Query Match 9.9%; Score 68.6; DB 5; Length 556; Best Local Similarity 48.3%; Pred. No. 8.6e-06; Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 265 CTCGCGTTTACTGTAAGTTTCTGCTTAGACCTCCGATCGCCATCGCATGCA 324  
DB 486 CCCCCCTCCCCCTTTTATATATTTTATTTTCTCCCCCCCCCTCCCATTAATA 427

QY 325 TTCTGCTCGATTCTCTTTTCTCGCGAATAATGCTTAATGTTCTGATTTG 384  
DB 426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 367

QY 385 AAGTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTTAGTAAGATA 444  
DB 366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307

QY 445 CCGCGCTTACTGTTTGTTCATTTGTGTCGTTTACGCTTAGCGGATCGCA 504  
DB 306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247

QY 505 GTATTTGACGTGAAAAATCTTCGTTTGTGTTTGTTCATTAATCGATGAT 564  
DB 246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187

QY 565 CTACCTTTGTGCTTGTGATGTTGTTTGTGAGCCATAGCGTTGTGTAATCT 624  
DB 186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127

QY 625 TCACGTTCAATGTTGATTTGAGATTTT 653  
DB 126 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 98

RESULT 5  
ABV42105/c  
ID ABV42105 standard; cDNA; 556 BP.  
XX ABV42105;  
AC  
XX  
DT 16-SEP-2002 (first entry)  
DE Human prostate expression marker cDNA 42096.  
XX  
XX Human prostate expression marker cDNA 42096.  
XX  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX phamacogenomic marker; gene; ss.  
XX Homo sapiens.  
XX OS  
XX WO200160860-A2.  
XX PD 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.  
PF 17-FEB-2000; 2000US-0183319P.  
XX 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0213314P.  
PR 18-JUL-2000; 2000US-0219077P.  
PR 13-DEC-2000; 2000US-0255281P.  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PI Schlegel R, Endege WO, Monahan JR;  
XX WPI; 2001-662795/76.  
DR  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1; Page 8441; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

SO Query Match 9.9%; Score 68.6; DB 5; Length 556; Best Local Similarity 48.3%; Pred. No. 8.6e-06; Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 265 CTCGCGTTTACTGTAAGTTTCTGCTTAGACCTCCGATCGCCATCGCATGCA 324  
DB 486 CCCCCCTCCCCCTTTTATATATTTTATTTTCTCCCCCCCCCTCCCATTAATA 427

QY 325 TTCTGCTCGATTCTCTTTTCTCGCGAATAATGCTTAATGTTCTGATTTG 384  
DB 426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 367

QY 385 AAGTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTTAGTAAGATA 444  
DB 366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307

QY 445 CCGCGCTTACTGTTTGTTCATTTGTGTCGTTTACGCTTAGCGGATCGCA 504  
DB 306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247

QY 505 GTATTTGACGTGAAAAATCTTCGTTTGTGTTTGTTCATTAATCGATGAT 564  
DB 246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187

QY 565 CTACCTTTGTGCTTGTGATGTTGTTTGTGAGCCATAGCGTTGTGTAATCT 624  
DB 186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127

QY 625 TCACGTTCAATGTTGATTTGAGATTTT 653  
DB 126 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 98

RESULT 6  
ABV43601/c  
ID ABV43601 standard; cDNA; 556 BP.

XX ABV43601;  
AC 16-SEP-2002 (first entry)  
XX  
DT Human prostate expression marker CDNA 43592.  
XX  
DE Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.  
XX Homo sapiens.  
XX WO200160860-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US005171.  
XX  
XX 17-FEB-2000; 2000US-0183319P.  
XX 16-MAR-2000; 2000US-0189862P.  
XX 25-MAY-2000; 2000US-0207454P.  
XX 09-JUN-2000; 2000US-0211314P.  
XX 18-JUL-2000; 2000US-0219007P.  
XX 13-DEC-2000; 2000US-0255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
XX Claim 1; Page 8681-8682; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX specification or its complement. (I) is useful for: (a) assessing whether  
XX a patient is afflicted with prostate cancer; (b) monitoring the efficacy  
XX of a test compound to inhibit prostate cancer in a patient; (c) assessing  
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;  
XX (d) selecting a composition for inhibiting prostate cancer in a patient;  
XX (e) assessing the prostate cell carcinogenic potential of a compound; (g)  
XX determining whether prostate cancer has metastasized in a patient; (h)  
XX assessing the aggressiveness or indolence of prostate cancer in a patient  
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
XX Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;  
XX  
XX Query Match 9.9%; Score 68.6; DB 5; Length 556;  
XX Best Local Similarity 48.3%; Pred. No. 8.6e-06;  
XX Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;  
XX  
QY 255 CTCGCTGCTTACTGCTAGTCTTCTGCTAGACCTCCGATGCGCTACCGCATGCA 324  
DB 486 CCCCCCTCCCTCCCTTTTAAATATTTTCTTCTCCCTCCCTCCCTCCCTCCCTAAATA 427  
QY 325 TTCGCTGCTGATCTCTTCTCTGCTGGAATAATGCCCTAATGTTCTGATTTGG 384  
DB 426 TTTTCTTTTCTTTTCTTTTCTTTTAAATAAATAAACCCCTTTTCTTTTCTTTT 367  
QY 385 AAGGTTTGTGCTATGGCTTACTTTTCCCTAATATTTAATAGTTCTTAGGTAAGATA 444  
DB 366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 307  
QY 445 CCGTGGCTTACTGTTTGTGCTATTTGTGCTTACCGGTTAGTCGCTGATCGGA 504  
DB 306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 247  
QY 505 GATTTGACTGTGAAATATCTGTTTCTTTGTTTCTTCAATTAATGAGATGAT 564

DB 246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 187  
QY 565 CTACCTTTTGCTGCTGATGTTGTTTGTGACCTAGCCCTGCTGCTTATAACT 624  
DB 186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 127  
QY 625 TCACGCTGATGCTGATTTTGAATTTT 653  
DB 126 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 98  
XX  
XX RESULT 7  
XX AAS45347  
XX ID AAS45347 strand; DNA; 9539 BP.  
XX  
XX AAS45347;  
XX  
XX 18-DEC-2001 (first entry)  
XX  
XX Chemically pretreated complementary DNA associated with cell cycle #26.  
XX  
XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
XX human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
XX graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
XX arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
XX immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
XX PCR primer.  
XX  
XX Homo sapiens.  
XX  
XX WO200168911-A2.  
XX  
XX 20-SEP-2001.  
XX  
XX 15-MAR-2001; 2001WO-EP002945.  
XX  
XX 15-MAR-2000; 2000DE-01013847.  
XX 06-APR-2000; 2000DE-01019058.  
XX 07-APR-2000; 2000DE-01019173.  
XX 30-JUN-2000; 2000DE-01032529.  
XX 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-602751/68.  
XX  
XX Designing primers and probes for analyzing diseases associated with  
XX cytosine methylation state e.g. arthritis, cancer, aging  
XX arteriosclerosis comprising fragments of chemically modified genes  
XX associated with cell cycle.  
XX  
XX Claim 1; SEQ ID NO 52; 28pp; English.  
XX  
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
XX molecules associated with the cell cycle and specific PCR primers of the  
XX invention. The sequences are useful for detecting the methylation state  
XX of all CpG dinucleotides in a sequence and therefore for analysing  
XX associated diseases. By analysing cytosine methylations in the pretreated  
XX DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
XX of existing diseases or the predisposition to specific diseases can be  
XX ascertained. The parameters may be compared to another set of genetic  
XX and/or epigenetic parameters, the differences serving as basis for  
XX diagnosis and/or prognosis events which are disadvantageous to patients.  
XX The sequences of the invention are useful for the diagnosis and therapy  
XX of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
XX aging, glomerular disease, Lewy body disease, arthritis,  
XX arteriosclerosis, solid tumours and cancers  
XX  
XX Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;





```
RESULT 9
AA33181/C
ID AAX33181 standard; DNA; 6644 BP.
XX
AC AAX33181;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRc-lres-bsr.
XX
KM Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KM autoimmune disease; graft rejection reaction; inflammation;
KM inflammatory disease; ss.
XX
OS Synthetic.
OS Cowpox virus.
XX
PN WO913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 1; Page 38-41; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRc-lres-bsr, which contains the cowpox virus bsr gene, and is used in an
CC example from the present invention
XX
SQ Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 U; 0 Other;
XX
Query Match 9.5%; Score 66; DB 2; Length 6644;
Best Local Similarity 45.4%; Pred. No. 5.6e-05;
Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0;
OY 157 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGTCAGTTATTTCTTCTCC 216
DB 4235 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 4176
OY 217 GGAGCGTCGACCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 276
DB 4175 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 4116
OY 277 CTCGTAGATTCTTGCCTTAGAGCCTCGATCGCTCAGCCGATGATTTCTGTCTGCA 336
```

```
DB 4115 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 4056
OY 337 TTTCTCTTTTCTTGGCGGAAAAATGCGCCAAATGCTCGATTTCGAGCTTTTG 396
DB 4055 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3996
OY 397 CTATGGTCTACTTTTCCCTAATTTTATAGTCTTAGAGTAACGATACCTGCTTAC 456
DB 3995 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3936
OY 457 TCTTTTGTTCATTTTGTGCTTTCACCGTTAGTCGCTGAGATATTGACTGT 516
DB 3935 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3876
OY 517 GAAAAATCCTTGCTTTTGTGTTTGTCTTCAATAAATGAGATGATCTTACTTTG 576
DB 3875 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3816
OY 577 CTTCATGTTGTTTCTTTGAGCCCTAAGCGTTGTGCTTGTATTAACCTGACTGAT 636
DB 3815 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3756
OY 637 GTGATTTTGAGATTGTGTAGTGA CTGTGGCTTTCTTTGT 678
DB 3755 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3714

RESULT 10
AA33182/C
ID AAX33182 standard; DNA; 7372 BP.
XX
AC AAX33182;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRc-Bcl-x1-bsr.
XX
KM Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KM autoimmune disease; graft rejection reaction; inflammation;
KM inflammatory disease; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 2; Page 41-45; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
```

therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the base sequence of the plasmid PRX-Bcl-xl-bst, which contains the human Bcl-xl gene, and is used in an example from the present invention

Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 U; 0 Other;

Query Match 9.5%; Score 66; DB 2; Length 7372;  
Best Local Similarity 45.4%; Pred. No. 5.7e-05;  
Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

```
QY 157 TTTTGAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGTCAGTTTATTTCTTCCTCC 216
    |||||
DB 4963 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4904
QY 217 GGAGTCCGACCTACACTCTCAGCTCCGCGCTTAACTAGCTCTCCGCTGTTA 276
    |||||
DB 4903 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4844
QY 277 CTCTGAAGTTTCTGCTTAGAGCCCTCGATCGCTCAGCCGATCTGCTGCTGA 336
    |||||
DB 4843 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4784
QY 337 TTTCTCTTTTCTGCTCGGAAAAATGCCCCATATCTTCATTCGATTCGAAGTTTGTG 396
    |||||
DB 4783 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4724
QY 397 CTATGGGTACTTTTCTCCATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 456
    |||||
DB 4723 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4664
QY 457 TGTATTTGTCATTTTGTGCTTACCGCTTACGCTTACGCTGATTCGAGTTCGT 516
    |||||
DB 4663 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4604
QY 517 GAAAAATCTTCCTTTTGTGTTTGTTCATATAAATCGATACCTTCTG 576
    |||||
DB 4603 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4544
QY 577 CTTTGAGTTTGTATTTTGTAGCCTATGCGTTGTGCTTTATTAACCTTCACGTTGAT 636
    |||||
DB 4543 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4484
QY 637 GTGGATTTTGAAGTTTGTAGTACTGCGGTTCTTTGCT 678
    |||||
DB 4483 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4442
```

RESULT 11  
AAK33180/c  
ID AAK33180 standard; DNA; 7797 BP.

XX AAK33180;

XX 25-JUN-1999 (first entry)

XX Cowpox virus bsr full length gene sequence.

XX Cowpox virus: bsr; viral vector; expression; apoptosis; resistance; crmA;

XX bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;

XX auto-immune disease; graft rejection reaction; inflammation;

XX inflammatory disease; ss.

XX Cowpox virus.

XX

PN W09913073-A2.  
XX  
XX 18-MAR-1999.  
XX  
XX 07-SEP-1998; 98WO-JP004010.  
XX  
XX 08-SEP-1997; 97JP-00259235.  
XX  
XX (RRG-) RPR GENCELL ASIA PACIFIC INC.  
XX  
XX Hamada H;  
XX  
XX WPI, 1999-243728/20.  
XX  
XX New apoptosis-resistant virus-sensitive cell.  
XX  
XX Example 1; Page 34-38; 51pp; English.

The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction. therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the cowpox virus bsr gene which is used in an example from the present invention

Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 U; 0 Other;

Query Match 9.5%; Score 66; DB 2; Length 7797;  
Best Local Similarity 45.4%; Pred. No. 5.8e-05;  
Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

```
QY 157 TTTTGAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGTCAGTTTATTTCTTCCTCC 216
    |||||
DB 5388 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5329
QY 217 GGAGTCCGACCTACACTCTCAGCTCCGCGCTTAACTAGCTCTCCGCTGTTA 276
    |||||
DB 5328 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5269
QY 277 CTCTGAAGTTTCTGCTTAGAGCCCTCGATCGCTCAGCCGATCTGCTGCTGA 336
    |||||
DB 5268 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5209
QY 337 TTTCTCTTTTCTGCTCGGAAAAATGCCCCATATCTTCATTCGATTCGAAGTTTGTG 396
    |||||
DB 5208 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5149
QY 397 CTATGGGTACTTTTCTCCATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 456
    |||||
DB 5148 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5089
QY 457 TGTATTTGTCATTTTGTGCTTACCGCTTACGCTTACGCTGATTCGAGTTCGT 516
    |||||
DB 5088 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5029
QY 517 GAAAAATCTTCCTTTTGTGTTTGTTCATATAAATCGATACCTTCTG 576
    |||||
DB 5028 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4969
```



PT Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.

PS Claim 1; SEQ ID NO 1931, 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel  
CC diseases. The present sequence is a gene of the invention

XX Sequence 14006 BP; 3278 A; 155 C; 3257 G; 7313 T; 0 U; 3 Other;

Query Match 9.5%; Score 66; DB 6; Length 14006;  
Best Local Similarity 45.3%; Pred. No. 6.5e-05;  
Matches 240; Conservative 0; Mismatches 290; Indels 0; Gaps 0;

QY 157 TTTTGGAGCTTAATAGCTTCCGATTTGCTCTCTGCTGAGTTTATTTCTCTCC 216  
DB 1974 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2033  
QY 217 GGAGTCTGACTACTACTCTGCTGCGGCTTAACCTTACGTTCTCGCTGTTA 276  
DB 2034 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2093  
QY 277 CTCTGTAAGTTTTCGCTTACAGCTCCGAGTCCGCTACCGAGTCTGTGCTGA 336  
DB 2094 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2153  
QY 337 TTTTCTTTTCTCTGCTGGAATAATGCTTAAGTCTGATTTGGAAGTTTGTG 396  
DB 2154 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2213  
QY 397 CTATGGGTACTTTTCCCTAATTTTATTTAGTTCTTAAAGTACCTGCTTAC 456  
DB 2214 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2273  
QY 457 TGTATTGTTGATTTTGTGCTTACCGCTTACGCTGATGAGATTTGACTGT 516  
DB 2274 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2333  
QY 517 GAAAAATCCTGCTTTTGTGTTTGTTCATTAATGAGTATCTACCTTTGTG 576  
DB 2334 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2393  
QY 577 CTATGATTTGTTTATTTGAGCTTACGCTTGTGCTTGTATTAAGTCTGATGT 636  
DB 2394 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2453  
QY 637 GTGATTTTGAATTTGTAGTACGTTGGTCTTCTTGATGCTATAG 686  
DB 2454 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2503

RESULT 14  
ABO36997/c  
ID ABO36997 standard; DNA; 556 BP.

XX ABO36997;  
XX  
XX 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 23588.  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP010074.

XX 01-SEP-2000; 2000DE-01043826.

XX 05-SEP-2000; 2000DE-01044543.

XX (EPITG-) EPITGONOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C to uracil, then part of the genomic  
CC DNA contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABO13410-  
CC ABO34121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention

XX Sequence 556 BP; 505 A; 14 C; 13 G; 12 T; 0 U; 12 Other;

Query Match 9.5%; Score 65.8; DB 6; Length 556;  
Best Local Similarity 49.3%; Pred. No. 3.6e-05;  
Matches 172; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 325 TTTGAGCTGATTTCTCTTTTCTGCTGAAAAATGCTTAAGTCTGATTTG 384  
DB 461 TTTGCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 402  
QY 385 AAGTTTGTGCTAAGGTTACTTTTCCCTAATTTTATTTAGTTCTTAAAGATA 444  
DB 401 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 442  
QY 445 CCGGCTTACTGTTTGTTCATTTGTGCTTACCGCTTACGCTGATCGGA 504  
DB 341 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 282  
QY 505 GATTTGACTGTAATAATCTGCTTTTGTGTTTGTTCATTAATGAGATTTAT 564  
DB 281 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 222  
QY 565 CTACCTTTTGTGCTTGAATTTGTTTGTGACCTAAGCTTGTGCTGTTAACT 624  
DB 221 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 162  
QY 625 TCAAGTCAAGTGTGATTTTGAATTTGTAGTACGCTGAGGTTCT 673  
DB 161 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 113

RESULT 15  
ABQ36996  
ID ABQ36996 standard; DNA; 556 BP.  
XX  
AC ABQ36996;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 23587.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP010074.  
XX  
PR 01-SEP-2000; 2000DE-01043826.  
PR 05-SEP-2000; 2000DE-01044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI: 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ3410-  
CC ABQ3412 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX  
SQ Sequence 556 BP; 12 A; 13 C; 14 G; 505 T; 0 U; 12 Other;

Query Match 9.5%; Score 65.8; DB 6; Length 556;  
Best Local Similarity 49.3%; Pred. No. 3.6e-05;  
Matches 172; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 325 TTCGTGCTGCAATTCCTCTTCGCTGGAATAATGCCATAGTTCGATTCG 384  
DB 96 TTTCGTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTTATTTTTTTTTTTT 155  
QY 385 AAGTTTGGCTATGGGTTACTTTTCCCTATATTTATAGTTCTTAGGTAAGATA 444

DB 156 TTTTTTTTTTTTTTTTTTTTTTTTTTTTACGTTTTTTTTTCGTTTTTTTTTTTTT 215  
QY 445 CCGCGCTTACTGTTTTTGTTCATTTGTGNGCTTTCACCGTTTAGTCGATCGGA 504  
DB 216 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTTT 275  
QY 505 GATTTGACGTGAAAAATCCCTGTTTTTTGGTTTTGTTCAATAAATCGATTCAT 564  
DB 276 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTTT 335  
QY 565 CTACCTTTGCTGCTTGTGATGTTGTTTGTAGCCCTAGCGTTGTGCTGTATTA 624  
DB 336 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTTT 395  
QY 625 TCACGTTAATGATGATTTGAGATTTGATGATGATGATGATGATGATGATGAT 673  
DB 396 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTTT 444

Search completed: March 29, 2004, 17:14:35  
Job time : 249.442 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 14:52:55 ; Search time 2617.78 Seconds

(without alignments)  
11507.244 Million cell updates/sec

Title: US-10-030-829-1\_COPY\_1\_695

Perfect score: 695

Sequence: 1 gacaaacaacaataa.....gtcgcctatagttgtaaa 695

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_srb:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_ov:\*  
23: em\_ov:\*  
24: em\_ov:\*  
25: em\_ov:\*  
26: em\_ov:\*  
27: em\_ov:\*  
28: em\_ov:\*  
29: em\_ov:\*  
30: em\_ov:\*  
31: em\_ov:\*  
32: em\_ov:\*  
33: em\_ov:\*  
34: em\_ov:\*  
35: em\_ov:\*  
36: em\_ov:\*  
37: em\_ov:\*  
38: em\_ov:\*  
39: em\_ov:\*  
40: em\_ov:\*  
41: em\_ov:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	695	100.0	3275	6	AX078760
2	695	100.0	81365	8	AB025633
3	85.6	12.3	7218	6	166494
4	84	12.1	2162	8	BT002944
5	83.4	12.0	184535	2	CNS05TCJ
6	81.2	11.7	349980	6	AX344566
7	72.8	10.5	1407	8	AJ592026
8	70	10.1	1453	8	AJ592978
9	70	10.1	66993	2	AC138074
10	69.6	10.0	1434	8	AJ592058
11	69.2	10.0	3364	14	TSP418778
12	69	9.9	88932	2	AC022680
13	68.8	9.9	6666	6	AX346599
14	68.8	9.9	81120	2	AC022851
15	68.4	9.8	810	8	AJ592180
16	68.2	9.8	56857	2	AC021917
17	68	9.8	1184	8	AJ591982
18	67.6	9.7	151368	2	AC142664
19	67.4	9.7	155550	2	AC015830
20	67.4	9.7	196853	9	CNS01DUD
21	67	9.6	1228	8	AJ592059
22	67	9.6	60565	2	AC023852
23	67	9.6	127709	2	CNS05TCJ
24	67	9.6	202083	2	AC023833
25	66.8	9.6	164288	9	AC128949
26	66.6	9.6	749	9	HS336745
27	66.6	9.6	9539	6	AX277889
28	66.6	9.6	9539	6	AX323566
29	66.4	9.6	83440	2	AC024285
30	66.4	9.6	194180	2	AC069140
31	66	9.5	6644	6	E23356
32	66	9.5	7372	6	E23357
33	66	9.5	7797	6	E23355
34	66	9.5	7996	6	E23359
35	66	9.5	14006	6	AX346860
36	66	9.5	56693	2	AC084077
37	66	9.5	143585	2	AC013349
38	65.8	9.5	64767	2	AC102701
39	65.8	9.5	273807	2	AC025421
40	65.8	9.5	289973	2	AC135678
41	65.6	9.4	113515	6	AX347076
42	65.4	9.4	6079	6	AX345323
43	65.4	9.4	93791	2	AC138073
44	65.4	9.4	174384	2	AC009524
45	65.4	9.4	242885	2	AC097836

# ALIGNMENTS

RESULT 1  
AX078760 3275 bp DNA linear PAT 22-FEB-2001  
LOCUS AX078760 Sequence 1 from Patent WO0105951.  
ACCESSION AX078760  
VERSION AX078760.1 GI:13158379  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE  
1 Beclin,C., Elmayan,T. and Vaucheret,H.  
Novel eg3 plant gene and use thereof

JOURNAL Patent: WO 0105951-A 1 25-JAN-2001;  
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE  
 AGRONOMIQUE (FR)  
 FEATURES Location/Qualifiers  
 source 1..3275  
 /organism="Arabidopsis thaliana"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3702"  
 primer\_bind 693..715  
 /note="p356AD"  
 primer\_bind complement(2926..2952)  
 /note="p356Y"  
 ORIGIN  
 Query Match 100.0%; Score 695; DB 6; Length 3275;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-130; Mismatches 0; Gaps 0;  
 Matches 695; Conservative 0; Indels 0; Gaps 0;  
 1 GACAAACAAACAAATTAAGCAAGTATGTCGACCAATTAATTAATGCGAACA 60  
 1 GACAAACAAACAAATTAAGCAAGTATGTCGACCAATTAATTAATGCGAACA 60  
 61 TTAAATTAGGAG 120  
 61 TTAAATTAGGAG 120  
 121 GAAATTTGAGTCCAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 121 GAAATTTGAGTCCAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 181 ATTGTCCTCTCTGTCGATGATTTATTTCTCTCCGAGTCGACCTACTCTCAC 240  
 181 ATTGTCCTCTCTGTCGATGATTTATTTCTCTCCGAGTCGACCTACTCTCAC 240  
 241 TCTCCGGGCTTAACTAGGTTTCCGCTGTTACTCTGTAAGTTTCTGCTTAA 300  
 241 TCTCCGGGCTTAACTAGGTTTCCGCTGTTACTCTGTAAGTTTCTGCTTAA 300  
 301 CCTCCGATCGCTCAACCGATGATTCGTCGATTTCTTTTCTGCTGAGAAA 360  
 301 CCTCCGATCGCTCAACCGATGATTCGTCGATTTCTTTTCTGCTGAGAAA 360  
 361 ATGGCCCTAATGTTCTGATTTGCAAGTTTGGCGATAGGCTACTTTTCCCTA 420  
 361 ATGGCCCTAATGTTCTGATTTGCAAGTTTGGCGATAGGCTACTTTTCCCTA 420  
 421 TTTTATAGTTCTTAGTAAGATACCTGCTTACTGTTTGTTCATTTTGTGTCT 480  
 421 TTTTATAGTTCTTAGTAAGATACCTGCTTACTGTTTGTTCATTTTGTGTCT 480  
 481 TTACACGTTTAACTGCTGATCGAGATTTGACTGTGAAAAATCCTGTTTGT 540  
 481 TTACACGTTTAACTGCTGATCGAGATTTGACTGTGAAAAATCCTGTTTGT 540  
 541 TTGTTTCATTAATGAGATGATGATCACTGTTTGTGCTTGAAGTTTGTGAG 600  
 541 TTGTTTCATTAATGAGATGATGATCACTGTTTGTGCTTGAAGTTTGTGAG 600  
 601 ATGCGTTTGGCTGTTAATCTTACGCTTCAATGTCGATTTGAGATTTGTAG 660  
 601 ATGCGTTTGGCTGTTAATCTTACGCTTCAATGTCGATTTGAGATTTGTAG 660  
 661 ACTGTGGGTTTCTTTGGTGGCTTAAAGTTTATAA 695  
 661 ACTGTGGGTTTCTTTGGTGGCTTAAAGTTTATAA 695  
 RESULT 2  
 AB025633 81365 bp DNA linear PLN 27-DEC-2000  
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MOM1.  
 DEFINITION  
 ACCESSION AB025633 BA000015  
 VERSION AB025633.2 GI:10178221

KEYWORDS  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
 1 (sites)  
 Saito, S., Nakamura, Y., Kaneko, T., Katoh, T., Asanizu, E., Kotani, H.,  
 and Tabata, S.  
 Structural analysis of Arabidopsis thaliana chromosome 5. X.  
 Sequence features of the regions of 3,076,755 bp covered by sixty  
 P1 and PAC clones  
 DNA Res. 7 (1), 31-63 (2000)  
 JOURNAL MEDLINE  
 PUBMED 10718197  
 2 (bases 1 to 81365)  
 Nakamura, Y.  
 Direct Submission  
 Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research  
 Institute, Department of Plant Gene Research, 1532-3, Yana,  
 Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,  
 Tel: 81-438-52-3935, Fax: 81-438-52-3934)  
 On Sep 15, 2000 this sequence version replaced gi:4589439.  
 Address for correspondence: kaos@kazusa.or.jp  
 For the latest information on annotation of this clone, please see  
 http://www.kazusa.or.jp/kaos/cgi-bin/agd.graph.cgi?c=MOM1  
 Genes with similarity to proteins in the databases are described in  
 'product' or 'note' qualifiers. Genes that have no significant  
 protein similarity are described as 'unknown protein'.  
 The software programs used to predict genes include: Gail  
 (Informatics Group, Oak Ridge National Laboratory,  
 http://compbio.ornl.gov/Grail-1.3/),  
 GENSCAN (Chris Burge, MIT, http://CBR-081.mit.edu/GENSCAN.html),  
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of  
 Denmark, http://www.cbs.dtu.dk/services/netgene2/) and  
 SplicePredictor (Volker Brendel, Stanford University,  
 http://gremlin.zool.iastate.edu/cgi-bin/sp.cgi).  
 Genes encoding tRNAs are predicted by tRNAscan-SE  
 (Sean Eddy, Washington University School of Medicine, St. Louis,  
 http://genome.wustl.edu/eddy/tRNAscan-SE/).  
 This sequence may not be the entire insert of this clone. It may be  
 shorter because we remove overlaps between neighboring submissions.  
 The 5' clone is K19M13 and the 3' clone is MOM1.  
 Location/Qualifiers  
 1..81365  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 /clone="MOM1"  
 /clone\_id="Mitsui P1"  
 join(3184..3403,3474..3595,3681..3723,3777..3898,  
 4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)  
 /note="gb|AA01580.1  
 gene\_id:MOM1.24  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BA097251.1"  
 /db\_xref="GI:8809710"  
 /translacion="MAYOFGTSVDSRAHKKNDTPKRVQFLEPISKTVDNAONPS  
 YGSAFDDPSSENPLSPVEEPSSSTSEGGDDPLFGIEDLQISGEPYGRHEOAG  
 YSINGTSCNFEWVCHLEDGSVNYIDSSERRSNAAKQPNYLVADDDYLAIYV  
 QLPDDRKRGKRFPAFKYDLTVLIDVHTLTFSDCFVFEFWMFALFICYPILSG  
 ACDDPMOSNIEKTLHGHASYSKVLAVGFVDIWEAALTSIKRGYSIKICISDLTIE  
 KFSASTYTIPIPGOPAEVLIIIGSDGSHSLRADNGSPDLGSRDEIVLTALFKRW  
 FPKXNSFYHFRCSKSRGL"  
 join(6317..6428,6508..6629,6714..6756,6841..6932,  
 7121..7169,7296..7466,7550..7663,7785..7882,8112..8181)  
 /note="gb|AA01580.1  
 gene\_id:MOM1.23  
 similar to unknown protein"

/codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97250.1"  
 /db\_xref="GI:8809709"  
 /translation="MQPQTSVYESPALNQADEFPIKSTVYVNTKNIFFPDGDADL  
 PALNLSISGPIYGHLEQAGSISINETSINFVWCHLEDGSVNYIDGAKPPYLVLT  
 ADVDGLAEVQPLDIDRNRKGLVAVFANDNRKIACHPEMOSNIKTLTGHAISYV  
 SLAIGFVHWEATLSIEREGYTIKCNNDLITKFSASTAVKIPFEKPAELVIGSD  
 GSEHCLVDNEMPDIRLCKEKEKRVLRVKKVYVEOKL"  
 complement(join(8642..8705,8794..8863,8978..9387,  
 9584..10347))  
 /note="gene\_id:MQM1.22  
 unknown protein"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97249.1"  
 /db\_xref="GI:8809708"  
 /translation="MSWMKGSQMTAFDKORQKQIESEVEGDPPPVSTVSNA  
 GVRGLRNRHNSPSEKFSVLLPSPRPALLENDCQNERGCGCRKPTLILPVS  
 HDLAFKLKEMNSWADNLLRDVLLSTEDPEMALAFKGVSGKEDDEPTIEGY  
 SSNRSREYRFEKTIVTSVYMAASRPFDAQKVDLENSDGSFLVNASDNEKPD  
 SELDSTIQROQSTPIEPWEDDYLISHKDKALVMSASNSHSAQNAQRDHA  
 KQSDKREDMLAEKLNABAKKIIGITNKDNDIKLHLGHATAVQADLRDM  
 IEHFIVNRSVSPNRRSKNALASASQEPFGRLEDEGMHCQRTSRELNSLOVITG  
 IGHSRQASLPLAVKTFPFDNRYRPFETRPGVITVPRKFRHS"  
 complement(12417..13424)  
 /note="contains similarity to unknown protein  
 gb|AAB7018.1  
 gene\_id:MQM1.21"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97248.1"  
 /db\_xref="GI:8809707"  
 /translation="MATDSQBNOKLTPLKTRIALTYISTMTDQAQPDGTRRFR  
 LFRPAPRPKNVITSTSDPVQSDMLFRLTTPVSGDKIPVVFHGGGAFAS  
 PNAIPYDNCRRFAPKLPAYIVISVNYLAPHRPAQYDGFALAKYIENHGSILTA  
 NDLISRCFPAGDSAGNIANVNAIRICREPRSEFTAYIGLISIQPFGESETEAR  
 KQVAGPLVSPDRDTCWAKMLNDEAVNGPNAVDISGLDPEPTMVVAGFDL  
 KDMORSYEWMLKCGKATLIEYPMFHAIFYIPPELPEAGQLMRIKDPIVERASIS  
 A"  
 join(14922..14981,15072..15167,15563..15666,15755..15891,  
 16096..16178)  
 /note="gene\_id:MQM1.20"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="50S ribosomal protein L24"  
 /protein\_id="BAA97247.1"  
 /db\_xref="GI:8809706"  
 /translation="MGWKAELKLRHKLIRGDNVMIIRGDKETGTIRVIRSNR  
 VIEGKMLIKKIKGPDHEGGIETVEAPLHASVQVDPVTPGRPCVGVKGLDEGK  
 VRVARGTSGSLIPRELIKIRATPRTPAGSKDPTMEFVQOTYAKTKGKMPUL"  
 join(16823..16952,17143..17320,17412..17493,17675..17851,  
 17943..18110,18199..18390)  
 /note="gene\_id:MQM1.19"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="26S proteasome, non-ATPase regulatory subunit"  
 /protein\_id="BAA97246.1"  
 /db\_xref="GI:8809705"  
 /translation="MERLQRIFGAGGLGHASPDSPDLTDSBOVYISLALLMKHG  
 RACVPMVWMLGEPVDEYTVVDFAPQSGTVSVAVDHFEQTNLMDLKQSG  
 RPEMVVGVYSHSGFGCMLSGVDINTQSGEALNORAVAVVPIOSVKGKVIDFR  
 SINPOTIMGOEPRQTTNSGHLNKPSIQALLHGLNHHVYSIAINRKNLEKMLR  
 LHKKKTDLGLTLPDTHSKTNEQTVQEMLSLAKTKAVQEEDELSPEKLAIVNGR  
 QDKKHLDEHVSMLMSNIVQTLGTMLDIVP"  
 complement(join(18965..19043,19240..19298,19419..19536,  
 19666..19739,19908..19961,20387..20530))  
 /note="contains similarity to unknown protein  
 emb|CAA22897.1  
 gene\_id:MQM1.18"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97245.1"

/db\_xref="GI:8809704"  
 /translation="MDKNOAIEKMKLVGMEVEDEQQAADSESSISEMEDLNRCAL  
 TTRORFYFALICLSAGLITCLLMLVFNFNKGIPTTLNMAIGSTAFILIGROV  
 TMDLDPARITATLYLASSITIALFCALYVNNKLLTLALTLLFGLTWLSLSIYPRK  
 TMSKIPMTCEPDEP"  
 join(22672..23634,23708..23999,24111..24355,24458..24624,  
 24715..24925)  
 /note="emb|CAB62356.1  
 gene\_id:MQM1.17  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97244.1"  
 /db\_xref="GI:8809703"  
 /translation="MSRAGPMKERNVGGVRYVEVQIVQIAGTFLASSODGGEW  
 EVISKKNKRPNTSGKTVVSONSPFRAGGQOQGRGVSNGRNVSGRNGNGC  
 IQANISGRGALSRYKDDNFPVPSRPLEGGMMQAGSGAQTAVQEPFVED  
 VDNASEBENDSDALDDSDLDADDDVDSQSHSGRKNKMPKFFGSLDSISIG  
 INBPQOMHPACONGPGLIDVNLHPLHARTKGRVYKLRHLEAVELEKOLMG  
 ASVIPGEITVGOKGLGEDEKQVETWMPVITIMTRLDKDDNDKTMGMNOBILAYF  
 DKREALRARSYGPQCHRGMSVUMPESSAGYIEARLHRELAEMGLDIRYAKGKSM  
 PSCGVROLVGLATKDDLDINSHSGKTRKLELSYQEMVRELRQISEDNQOLNY  
 FKRLKQNKHAFVLEESLEIMSKEKTRLEADRIVQRKQMOHEONREMDADRF  
 MDSIKQIHERRDAAKEFEMLOQOERAKVQOOQOINPESNDPCRRAEVSFIEF  
 OEKEMEFEVEREMLIKQOEKMEKMKRHEEIPLEKEFDALBQMLKYGHLNED  
 D"  
 join(25614..25838,26061..26307,26324..26439,26516..26610,  
 26710..26813,27522..27765,27943..28036,28115..28305,  
 28385..28604,28688..28855,28944..29092)  
 /note="gb|AAC97420.1  
 gene\_id:MQM1.16  
 strong similarity to unknown protein"  
 /codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97243.1"  
 /db\_xref="GI:8809702"  
 /translation="MAPAGQTAAVAABAAGADGAPPOQRGFGSTISGIVRAVFWYF  
 ASKEFPKQPMPTAPSOQMLTFKSGSLDMWFYLSBEKENDRGNDAALYHETN  
 IYAVWTPESIRPKSLTYVDSFTLQNGSLYATIPRSGSPFIDPDREYQPLNSER  
 THAVATYFPKQKNKKSLIGSKPDDESPEVEKQDKSDKEEYVPMVLSIMKEN  
 VTINLVDDFTQFADLVEPTGNYLPYIIFNEFWMLRDKETPNEVYSELPLMLEISP  
 ISMMKQLPQOVQOSFQOMORSYSGMLDGBDEKFLBENPYLLGITMFSMLHSVF  
 DFLAFNDILOFMNKNKSMESGLSAKSVLVNLCQFVILFLDNDTMSILASGSGVC  
 IEFWKIGKAMRIEVRSGMIPRLRPHDRSVYAKRKEVDDIAIKFLSYVLLLVGL  
 SIYSIAVERHSWYSLISLSTSCYVMFGIMCPQLFITYKLSVAHLEPWRQTYKF  
 LNTIIDLFAFYIKMPLHRLSVFRDVLFLILLYORWYVPVQKIRNENRFGGSDST  
 AEKKLTTERKEEEDKKTN"  
 complement(join(29439..29576,29661..29885,29985..30268,  
 30341..30493,30579..30722,30956..31484))  
 /note="gene\_id:MQM1.15"

Query Match 100.0%; Score 695; DB 8; Length 81365;  
 Best Local Similarity 100.0%; Pred. No. 7,8e-131;  
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GACAAACAAACAAATTAAGCAGTCATGCTGTAGCAATAATTAATGCGGAAACAA	60
DB	21977	GACAAACAAACAAATTAAGCAGTCATGCTGTAGCAATAATTAATGCGGAAACAA	22036
QY	61	TTAAGTTAAGCGAAAAAGAAAAAGAGTCAAAAAATGAAACAAATCAATGAAAT	120
DB	22037	TTAAGTTAAGCGAAAAAGAAAAAGAGTCAAAAAATGAAACAAATCAATGAAAT	22096
QY	121	GAAAAATTTGAGATCGCAATCGGAAAAAGCAGCCGTTTGAAGCTTAATAGCTTC	180
DB	22097	GAAAAATTTGAGATCGCAATCGGAAAAAGCAGCCGTTTGAAGCTTAATAGCTTC	22156
QY	181	ATTGTCCTCTTCTCGAGATTATTTCTCTCGAGATCTGACTACATCTCTCAC	240
DB	22157	ATTGTCCTCTTCTCGAGATTATTTCTCTCGAGATCTGACTACATCTCTCAC	22216
QY	241	TCTCGGCGCTTTAACTGATTCGCTGCTTGAATCTGTAAGTTTCTGCTTGAAG	300

Db	22217	TCGCCGCGCTTTAACTTAAGTTCTCCGCGTTTACTCGTAAGTTTCTGCGCTAAG	22276
QY	301	CCTCCGATCGCCTCACCGCAGCATCTGTGCTCGATTTCTCTTTTCTTCGCTGGAAAA	360
Db	22277	CCTCCGATCGCCTCACCGCAGCATCTGTGCTCGATTTCTCTTTTCTTCGCTGGAAAA	22336
QY	361	ATTGCCCTAATGCTCTGCAATTGCAAGTTTGTGCTAAGGTTACTTTTTCCCTATA	420
Db	22337	ATTGCCCTAATGCTCTGCAATTGCAAGTTTGTGCTAAGGTTACTTTTTCCCTATA	22396
QY	421	TTTTAATGTTCTTAAGTAACGATACCTGCGTTTACTGTTTTGTTCAATTTGTGTCT	480
Db	22397	TTTTAATGTTCTTAAGTAACGATACCTGCGCTTACTGTTTTGTTCAATTTGTGTCT	22456
QY	481	TTCAACCGTTTGTGCGTGAATGGAGATATTGAACGTGAAAAAATCTCGTTTTGGTT	540
Db	22457	TTCAACCGTTTGTGCGTGAATGGAGATATTGAACGTGAAAAAATCTCGTTTTGGTT	22516
QY	541	TTGTTTCATATAAATCGGATGATACCTAAGCTTTGTGCTTGAATGTTGTTTTGAGCCT	600
Db	22517	TTGTTTCATATAAATCGGATGATACCTAAGCTTTGTGCTTGAATGTTGTTTTGAGCCT	22576
QY	601	ATGCGTTGTGGCTGTATTAATACTCAAGTTGATGTGCAATTTTGAGATTTGGATGTG	660
Db	22577	ATGCGTTGTGGCTGTATTAATACTCAAGTTGATGTGCAATTTTGAGATTTGGATGTG	22636
QY	661	ACTGTGGGTTTCTTTGCTGCGCTAATAGTTGAAAA	695
Db	22637	ACTGTGGGTTTCTTTGCTGCGCTAATAGTTGAAAA	22671

```

RESULT 3
LOCUS      166494                      7218 bp    DNA
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION  166494
VERSION    166494.1    GI:2724471
KEYWORDS
SOURCE
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 7218)
AUTHORS   Dorrer,F., Schefflinger,F. and Falkner,F.Gunter.
TITLE      Recombinant fowlpox virus
JOURNAL    Patent: US 5670367-A 14 23-SEP-1997;
FEATURES   Location/Qualifiers
            1..7218
            /organism="Unknown"
            /mol_type="unassigned DNA"
ORIGIN

```

	Query Match	12.3%	Score 85.6 / DB 6	Length 7218	
	Best Local Similarity	6.4%	Pred. No. 1.7e-07		
	Matches 25 / Conservative	232 / Mismatches 131	Indels 0	Gaps 0	
Oy	174	CTTCCATATTGTCCTTTCTGTGCAGTTAATTTCTTCTCGGAGTCTTGACTCACTA	233		
		:::::	:::::	:::::	:::::
Db	1094	YY	1155		
Oy	234	CTCTCAGCTCTCGGGCCTTAAACTTAGCTTCCGCTGTTACTCTGAAGTTTTCTGC	293		
		:::::	:::::	:::::	:::::
Db	1154	YYY	1211		
Oy	294	CTTAGAGCTCCGATCGCTCACCGGATGATCTGTGCGAATTTCTTTTCTTCGCG	353		
		:::::	:::::	:::::	:::::
Db	1214	YYY	1271		
Oy	354	TGGAAAAATGCCCTATGTTCTCATTTGCAAGTTTTGTGCTATGGGTAATTTTTT	413		
		:::::	:::::	:::::	:::::
Db	1274	YYY	1331		
Oy	414	CCCTATATTTTATAGTCTTTAGGTAAAGATACCTGGGCTTACTGTTTTGTTCATTTTG	473		
		:::::	:::::	:::::	:::::

[illegible]

RESULT 4	BT002944	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
	BT002944	2162 bp	Arabidopsis thaliana clone RAFL14-93-K05 (R20243)	BT002944	BT002944.1	FLI cDNA.	Arabidopsis thaliana (chale crese)
		linear	unknown protein		GI:2754622		

Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
1 (bases 1 to 2152)  
REFERENCE  
AUTHORS

Onodera, C.S., Quach, H.L., Liang, C., Ioriama, M., Wong, S., Nishida, Y., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamitani, A., Kawai, T., Kim, C.J., Narisaka, M., Nuyens, M., Palm, C.J., Sakurai, T., Saitou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.  
Arabidopsis Full Length cDNA Clones

JOURNAL REFERENCE AUTHORS	Unpublished 2 (bases 1 to 2162) Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Tosiunt, M., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
TITLE	Direct Submission
JOURNAL	Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
COMMENT	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onders, C.S., Quach, H.L., Tang, C.C., Tortum, M., Wong, C., Wu, H.C., Yu, G., Yuan, S.S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M., Palm, C.J., Shih, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W., Becker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinzaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.

```
FEATURES
source
Location/Qualifiers
1..2162
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAF14-93-K05 (R20243)"
/ecotype="Columbia"
```

```

gene
5'UTR
CDS

/Note="This clone is in a modified plbluecript vector2
(lambda PS) as a BamHI/XhoI insert."
1..2162
/gene="At5g23570"
1..91
/gene="At5g23570"
92..1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AA02757.1"
/db_xref="GI:27754623"
/translation="MSSRAGMSSKKNVYGGVRRPEVLTVOGLACTRLASSDDGGE
EVISKKKKPKGNTSGKTVWSQNSNPFPAMGQQQGRSNVSGKNNVSGKNGR
IQAINISGRGALSRKYDNNFVAPPPVSPLEGGNNWQARGSGAQTAVQFFPYVEDD
VDNASEENDDSDALDDSDDDLASDIDSDVSQKSHSGRKKWPKFKPGSLDSIEK
INEDPORQWHPACQNGPGALIDWYNLHPLLAHARTGARVRLHLEAVLEKDLQMG
ASVIPCGETVIGMGKGLGDEKDYELVWPMVYINWTRLDKDNKKMIGMGOELLEVR
DKYELRLRHHYVGQHRGMSVMPRESSATYGLERLRHRELAEMGLDIRLWQGRSK
PSGGVRQLVGFLLATQDDIDFNQSHQGTFLKPELKSIOEWVLEKRISEDNQNLN
FKNKLSSKQNKAKVLESLLEISGLRTETADNRVLRQTMQHONNEADNRPEF
MDSIKQIHERDAAKEENFEMLQQOERAVVQQOQNIIPSSNDQCKKAAEEVSFIE
QEKEMEEFVEEREMLIKQEKEMEDMKRHHIEIFLDEKPEDALEQLMYGHGLHNEE
D"
1970..2162
/gene="At5g23570"
/misc_difference
2146
/gene="At5g23570"
/Note="not present in genomic sequence"

ORIGIN
Query Match 12.1%; Score 84; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 4.2e-07;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 199 AGTTATTTCTTCTCCGGAGTCTGACTACTACTCTCCGCGGCTTAACT 258
Db 1 AGTTATTTCTTCTCCGGAGTCTGACTACTACTCTCCGCGGCTTAACT 60
Oy 259 TACGTTCCGTCGTTACTCTGT 282
Db 61 TACGTTCCGTCGTTACTCTGT 84

RESULT 5
CNS05TCJ 184535 bp DNA linear HTG 06-JUN-2000
LOCUS Homo sapiens chromosome 14 clone R-691G13, WORKING DRAFT SEQUENCE,
8 ordered pieces.
ACCESSION ALJ55100
VERSION ALJ55100.2 GI:9213460
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 184535)
Genoscope.
Direct Submission
Submitted (06-JUN-2000)
On Jul 15, 2000 this sequence version replaced gi:8346752.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Contigs composition :
1386 bp contig from 1 to 1386
37234 bp contig from 1487 to 38720

```

```

1651 bp contig from 38821 to 40471
66410 bp contig from 40572 to 106981
16357 bp contig from 107082 to 123438
2460 bp contig from 123539 to 125998
20897 bp contig from 126099 to 146995
37440 bp contig from 147096 to 184535
-----
Overall quality chart :
Range      : bases
0           : 1377
1 - 9      : 1678
10 - 19    : 1569
20 - 29    : 3627
30 - 39    : 13450
40 - 49    : 13488
50 - 59    : 10047
60 - 69    : 22169
70 - 79    : 54644
80 - 89    : 43702
90 - 99    : 19284
-----
Percentage of bases with a quality value >= 40 : 88 %.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1       1386: contig of 1386 bp in length
*
*       1387       1486: gap of 100 bp
*
*       1487       38720: contig of 37234 bp in length
*
*       38721       38820: gap of 100 bp
*
*       38821       40472: contig of 1651 bp in length
*
*       40472       40571: gap of 100 bp
*
*       40572       106981: contig of 66410 bp in length
*
*       106982       107082: gap of 100 bp
*
*       107082       123438: contig of 16357 bp in length
*
*       123439       123538: gap of 100 bp
*
*       123539       125998: contig of 2460 bp in length
*
*       125999       126098: gap of 100 bp
*
*       126099       146995: contig of 20897 bp in length
*
*       146996       147095: gap of 100 bp
*
*       147096       184535: contig of 37440 bp in length.
*
Location/Qualifiers
1. 184535
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="14"
   /clone="R-691G13"
   /clone_1b="RPCT-11"
ORIGIN
Query Match      12.0%; Score 83.4; DB 2; Length 184535;
Best Local Similarity 32.3%; Pred. No. 3e-07;
Matches 218; Conservative 0; Mismatches 456; Indels 1; Gaps 1;

2 ACAACCAACAAATAATTAGCAAGTCACTGTCGTACCAATTAATTATCTGCGAACAAT 61
|||||
38591 AAAAAAAAAAAAAAAAAANANANANANANANANANANANANANANANANANANANAN 38650
|||||
62 TAAGTTAAGGAGAAAAAGAAAAAGGTACAAATGAAAAACAATACTGAATG 121
|||||
38651 AAAAAAAAAAAAAAAAAANANANANANANANANANANANANANANANANANANANAN 38710
|||||
122 AAATTGGAGTCAGAAATCGAAAAACGAGCCGTTTGAAGCTTAATAAGCTTCTCA 181
|||||
38711 AAAAAAAAAAAAAAAAAANANANANANANANANANANANANANANANANANANANAN 38770
|||||
182 TTGTCTCTTCCTGCGACGTTTATTTCTCTCCGCGAGTCTGACTACGTACTGACT 241

```

[illegible]

ORIGIN	
Query Match	11.7%; Score 81.2; DB 6; Length 349980;
Best Local Similarity	46.0%; Pred. No. 7.6e-07;
Matches 27;	Conservative 0; Mismatches 323; Indels 0; Gaps 0;
OY	80 AAAAAAAAAAGTACAAATAATGAAAAACAATCAACTGAATGAAAATTGGAGTCAGAA 139
Db	71449 AAAAAAAAAAAAAAAAAAAAAAAAAAATTAATATGAAAGTGTTATATTTTTAT 71508
OY	140 TCGGAAAAACGAGCGCTTTTAGCTTAATTAAGCTTCCTCATTTGTCTCTTTCGTCA 199
Db	71509 AAAGTAAATTGAGAATAGTTATCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 71568
OY	200 GTTATATTTTCCTCCCGAGTCCTGACCTACTACTGCATCCTCCGGCGCTTAAACTT 259
Db	71569 TTTTATT 71628
OY	260 ACCTGCCCGCTGCTTACCTGTAAGTTTCTGCTTAGAGCGCTCGATCGCTCACCGC 319
Db	71629 TTTTATT 71688
OY	320 ATGCATCTGTCGTGATTTCTCTTTCTCGCTGAAAAATGCCCTAATGTTTCGA 379
Db	71689 TTTTATT 71748
OY	380 TTTCGAAGTTTTGTCCTAATGGTTACTTTTCCCATTTTAAAGTCTTAGTAA 439
Db	71749 TTTTATATTT 71808
OY	440 CGATACCTGCTTACCTGTTGTTGTCATTTTGTCATTTGTCACCGTTAGTCGCTGA 499
Db	71809 TTTTATT 71868
OY	500 TCGAGATATTGACTGCAAAAAATCCTCGCTTTTGGTTTTGTTGTTATATAATCGGA 559
Db	71869 TTTTATT 71928
OY	560 TTGATCTACCTTTGCTTGATGTTGTTTTGTAGACCTAATCGTTGCGTTGTA 619
Db	71929 TTTTATT 71988
OY	620 TAACCTACGTTTCATGTCGATTTTGAGATTTTGTAGTACGTCGGTTCTTTGG 677
Db	71989 TTTTATTTGTATGTAG 72046
RESULT 7	
AJ592026	1407 bp DNA linear PLN 23-OCT-2003
LOCUS	Arabidopsis thaliana T-DNA flanking sequence, right border, clone 600D05
DEFINITION	
ACCESSION	AJ592026
VERSION	AJ592026.1 GI:37941650
KEYWORDS	right border; T-DNA flanking sequence.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE	1
AUTHORS	Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cnuad,C., Derose,R., Pelletier,G., Legnietec,L., Caboche,M. and Lecharny,A. T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites EMBO Rep. 3 (12), 1152-1157 (2002)
TITLE	
JOURNAL	
PUBMED	22363335
REFERENCE	2 (bases 1 to 1407)
AUTHORS	Balzergue,S.
TITLE	Direct Submision
JOURNAL	Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue



## COMMENT

Gaston Cremieux, 91057 Evry cedex, FRANCE  
 PCR was performed on DNA from transformants of *Arabidopsis thaliana* plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).

## FEATURES

## source

1. 1407  
 Location/Qualifiers

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/cultivar="Massilllewska"

/db\_xref="taxon:3702"

/clone="600D05"

/clone\_id="Arabidopsis thaliana T-DNA insertion lines"

1. 1407

/note="T-DNA flanking sequence"

## ORIGIN

## Query Match

Best Local Similarity 41.9%; Score 72.8; DB 8; Length 1407;

## Matches

224; Conservative 0; Mismatches 311; Indels 0; Gaps 0;

QY 157 TTTTAAAGCTTAATAGCTTCCATTTGCTCTCTCGTCACTTATTTCTCTCC 216  
 |||||  
 DB 703 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 762  
 QY 217 GGAGTCTGACTACTCTGACTCTCGGCGCTTAACCTAGCTTCCGCTTGA 276  
 |||||  
 DB 763 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 822  
 QY 277 CTCTGAGTTTCTGCTGAGAGCTCCGATCGCTCACCGAGCATTCGTGCTGCA 336  
 |||||  
 DB 823 TTTTATTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNT 882  
 QY 337 TTTCTCTTTTCTGCTGAGAAAATGCGCTTAATGCTCGATTCGAAAGTTTGTG 396  
 |||||  
 DB 883 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 942  
 QY 397 CTATGGTACTTTTCCCAATTTTATAGTCTTAGAGTACGATCCGCTTAC 456  
 |||||  
 DB 943 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTN 1002  
 QY 457 TGTATTTGTCATTTGTGCTTCAACGTTAGTGCATCGAGATTTGACGT 516  
 |||||  
 DB 1003 TTTTATTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 1062  
 QY 517 GAAAAATCCTGTTTGTGTTTGTTCATATAATCGATGATGATACCTTTGTG 576  
 |||||  
 DB 1063 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 1122  
 QY 577 CTATGAGTTTGTGTTTGAACCGATGCTTGTGCTTGTATTAACCTGATGAT 636  
 |||||  
 DB 1123 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 1182  
 QY 637 GTGATTTTGAATTTGTAGTACGTGGATTTCTTGTGCTGATGATGATGAT 691  
 |||||  
 DB 1183 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 1237

## RESULT 8

## AJ591978

## LOCUS

Arabidopsis thaliana T-DNA flanking sequence, right border, clone

## ACCESSION

AJ591978.1 GI:37941602

## KEYWORDS

right border; T-DNA flanking sequence.

## SOURCE

Arabidopsis thaliana (thale cress)

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE

1. Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Crund, C., Dehose, R., Pelletier, G., Lepoint, L., Caboche, M., and Leclercq, A.

## AUTHORS

T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites

## TITLE

EMBO Rep. 3 (12), 1152-1157 (2002)

## JOURNAL

2 (bases 1 to 1453)

## MEDLINE

22363535

## PUBMED

12446565

## REFERENCE

Balzergue, S.

## AUTHORS

Direct Submission

## TITLE

Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue

## JOURNAL

Gaston Cremieux, 91057 Evry cedex, FRANCE

## COMMENT

PCR was performed on DNA from transformants of *Arabidopsis thaliana* plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).

## FEATURES

Location/Qualifiers

## source

1. 1453

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/cultivar="Massilllewska"

/db\_xref="taxon:3702"

/clone="598D01"

/clone\_id="Arabidopsis thaliana T-DNA insertion lines"

1. 1453

/note="T-DNA flanking sequence"

## ORIGIN

## Query Match

Best Local Similarity 43.4%; Score 70; DB 8; Length 1453;

## Matches

232; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

QY 157 TTTTAAAGCTTAATAGCTTCCATTTGCTCTCTCGTCACTTATTTCTCTCC 216  
 |||||  
 DB 331 TTTTATTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNT 390  
 QY 217 GGAGTCTGACTACTCTGACTCTCGGCGCTTAACCTAGCTTCCGCTTGA 276  
 |||||  
 DB 391 TTTTATTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNT 450  
 QY 277 CTCTGAGTTTCTGCTGAGAGCTCCGATCGCTCACCGAGCATTCGTGCTGCA 336  
 |||||  
 DB 451 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 510  
 QY 337 TTTCTCTTTTCTGCTGAGAAAATGCGCTTAATGCTCGATTCGAAAGTTTGTG 396  
 |||||  
 DB 511 TTTTATTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNT 570  
 QY 577 CTATGAGTTTGTGTTTGAACCGATGCTTGTGCTTGTATTAACCTGATGAT 630  
 |||||  
 DB 631 TTTTATTTTNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 690  
 QY 517 GAAAAATCCTGTTTGTGTTTGTTCATATAATCGATGATGATGATGATGAT 576



[illegible]

1	1117	1116: contig of 1116 bp in length
*	1217	1216: gap of 100 bp
*	2337	2336: contig of 1120 bp in length
*	2337	2336: gap of 100 bp in length
*	2437	2436: contig of 1125 bp in length
*	3562	3561: gap of 100 bp
*	3662	4781: contig of 1120 bp in length
*	4782	4881: gap of 100 bp
*	4882	6001: contig of 1120 bp in length
*	6002	6101: gap of 100 bp
*	6102	7244: contig of 1143 bp in length
*	7245	7344: gap of 100 bp
*	7445	8458: contig of 1114 bp in length
*	8459	8558: gap of 100 bp
*	8559	9640: contig of 1082 bp in length
*	9641	9740: gap of 100 bp
*	9741	10846: contig of 1106 bp in length
*	10847	10946: gap of 100 bp
*	10947	12103: contig of 1157 bp in length
*	12104	12203: gap of 100 bp
*	12204	13320: contig of 1117 bp in length
*	13321	13420: gap of 100 bp
*	13421	14520: contig of 1100 bp in length
*	14521	14620: gap of 100 bp
*	14621	15773: contig of 1153 bp in length
*	15774	15873: gap of 100 bp
*	15874	16991: contig of 1118 bp in length
*	16992	17091: gap of 100 bp
*	17092	18240: contig of 1119 bp in length
*	18241	18340: gap of 100 bp
*	18341	19476: contig of 1136 bp in length
*	19477	19576: gap of 100 bp
*	19577	20721: contig of 1145 bp in length
*	20722	20821: gap of 100 bp
*	20822	21942: contig of 1121 bp in length
*	21943	22042: gap of 100 bp
*	22043	23141: contig of 1099 bp in length
*	23142	23241: gap of 100 bp
*	23242	24372: contig of 1131 bp in length
*	24373	24472: gap of 100 bp
*	24473	25627: contig of 1135 bp in length
*	25628	25727: gap of 100 bp
*	25728	26688: contig of 1161 bp in length
*	26689	26988: gap of 100 bp
*	26989	28030: contig of 1042 bp in length
*	28031	28130: gap of 100 bp
*	28131	29303: contig of 1173 bp in length
*	29304	29403: gap of 100 bp
*	29404	30554: contig of 1151 bp in length
*	30555	30654: gap of 100 bp
*	30655	31751: contig of 1097 bp in length
*	31752	31851: gap of 100 bp
*	31852	32666: contig of 1115 bp in length
*	32667	33066: gap of 100 bp
*	33067	34113: contig of 1047 bp in length
*	34114	34213: gap of 100 bp
*	34214	35329: contig of 1116 bp in length
*	35330	35429: gap of 100 bp
*	35430	36574: contig of 1145 bp in length
*	36575	36674: gap of 100 bp
*	36675	37828: contig of 1154 bp in length
*	37829	37928: gap of 100 bp
*	37929	39063: contig of 1135 bp in length
*	39064	39163: gap of 100 bp
*	39164	40289: contig of 1126 bp in length
*	40290	40389: gap of 100 bp
*	40390	41471: contig of 1082 bp in length
*	41472	41571: gap of 100 bp
*	41572	42654: contig of 1063 bp in length
*	42655	42754: gap of 100 bp
*	42755	43904: contig of 1170 bp in length
*	43905	44004: gap of 100 bp
*	44005	45128: contig of 1124 bp in length

*	45129	45228: gap of 100 bp	
*	45229	44538: contig of 1130 bp in length	
*	46359	44458: gap of 100 bp	
*	46439	47595: contig of 1137 bp in length	
*	47596	47695: gap of 100 bp	
*	47696	48785: contig of 1090 bp in length	
*	48786	48885: gap of 100 bp	
*	48886	49960: contig of 1075 bp in length	
*	49961	50060: gap of 100 bp	
*	50061	51132: contig of 1072 bp in length	
*	51133	51232: gap of 100 bp	
*	51233	52356: contig of 1124 bp in length	
*	52357	52456: gap of 100 bp	
*	52457	53556: contig of 1100 bp in length	
*	53557	53566: gap of 100 bp	
*	53657	54782: contig of 1126 bp in length	
*	54783	54882: gap of 100 bp	
*	54883	56008: contig of 1126 bp in length	
*	56009	56108: gap of 100 bp	
*	56109	57307: contig of 1199 bp in length	
*	57308	57407: gap of 100 bp	
*	57408	58529: contig of 1122 bp in length	
*	58530	58629: gap of 100 bp	
*	58630	59799: contig of 1170 bp in length	
*	59800	59899: gap of 100 bp	
*	59900	60983: contig of 1084 bp in length	
*	60984	61083: gap of 100 bp	
*	61084	62193: contig of 1110 bp in length	
*	62194	62293: gap of 100 bp	
*	62294	63397: contig of 1104 bp in length	
*	63398	63497: gap of 100 bp	
*	63498	64574: contig of 1077 bp in length	
*	64575	64674: gap of 100 bp	
*	64675	65800: contig of 1126 bp in length	
*	65801	65900: gap of 100 bp	
*	65901	66993: contig of 1093 bp in length	

```

FEATURES
source
location/Qualifiers
1..66993
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p"
/clone="RP13-907D8"
/clone_1b="RP13-13 Human Female BAC"

```

Query Match	10.1%	Score 70;	DB 2;	Length 66993;
Best Local Similarity	41.2%;	Pred. No. 0.00018;		
Matches 214;	Conservative	0;	Mismatches 305;	Indels 0;
			Gaps	0;

Qy	157	TTTTAGAGCTTAATACCTCCATCTTGCTCTCTGCTAGTTATATTTCTCTCTCC	216
Db	56776	TT	56835
Qy	217	GGAGTCCCTGACTACACTCTCACTCCGGCGCTTAAACTAGCTCTCGTCTGTA	276
Db	56836	TTNNNTNNNTNNNTTT	56895
Qy	277	CTCTGAAGTTTCTGCTTAGAGCTCGATGACCTCACCGATGCAATCTGTGCTGA	336
Db	56896	TTTTTNNNNNNNTTTNT	56955
Qy	337	TTTCTCTTTTCTCGCTGMAAAATGCCCTAATGCTCGATTTGAAAGTTTTGTG	396
Db	56956	TTTTTTTTTTTTTTTTTTTTNNNTTTTTTTTTTTTTTTTTTTTTNNNTTTTTTTT	57015
Qy	397	CTATGGGTACTTTTCCCTAATTTATAGTCTTAGTAAAGATACCTGCTCTAC	456
Db	57016	TTTTTTTTTTTTTTTTNNNTTTTTTTTTTTTTTTTTTTTTNNNNNNNTTNTTN	57075
Qy	457	TGTTTTGTTCATTTGTGTGCTTTCACCGTTAGTCGCTGATCGAGATTTGACTGT	516

[illegible]

RESULT 10	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AJ592058			AJ592058				
		1434 bp DNA linear					
		Arabidopsis thaliana T-DNA flanking sequence, right border, clone 602D05.					
			AJ592058				
			AJ592058.1	GI:37941682			
					right border; T-DNA flanking sequence.		
					Arabidopsis thaliana (thale cress)		
					Arabidopsis thaliana		

## REFERENCE AUTHORS

Chavvin, S., Bechtold, N., Cnuad, C., Derose, R., Peltelier, G.,  
Leplintec, L., Caboche, M. and Lecharny, A.  
T-DNA integration into the Arabidopsis genome depends on sequences  
of pre-insertion sites  
EMBO Rep. 3 (12), 1152-1157 (2002)

PUBMED	12446565
REFERENCE	2 (bases 1 to 1434)

**TITLE** Direct Submission  
**JOURNAL** Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue

**COMMENT**

plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.inbio.gen.fr>).

## FEATURES

```

misc_feature
      /organism="Arabidopsis thaliana"
      /mol_type="genomic DNA"
      /cultivar="Massilllewskija"
      /db_xref="taxon:3702"
      /clone="602D05"
      /clone_id="Arabidopsis thaliana T-DNA insertion lines"
1..1434
      /note="T-DNA flanking sequence
right border"

```

## ORIGIN

	Query Match	10.0%	Score 69.6;	DB 8;	Length 1433;
	Best Local Similarity	43.7%	Pred. No. 0.00036;		
	Matches 234;	Conservative	0;	Mismatches 301;	Indels 0; Gaps 0
OY	TTTTAGACCTTAATAGCCTTCATTGTCCTTCTTGCGAGTTATTCTCTCCGC	216			
b	TTTTTTTNTNNNTTT	562			



```

DB                2225 GTTATTGTGTTGTT 2239

RESULT 12
AC022680
LOCUS          Homo sapiens chromosome 8 clone RP11-511E14 map 8, LOW-PASS
DEFINITION
ACCESSION     AC022680
VERSION       AC022680.2 GI:9143927
KEYWORDS      HTG; HTGS PHASE0.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 88932)
AUTHORS       Birren,B., Linton,L., Nussbaum,C. and Lander,E.
TITLE         Homo sapiens chromosome 8, clone RP11-511E14
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 88932)
AUTHORS       Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Baldwin,J., Barna,N., Beckler,J., Beda,F.,
              Boguski,L., Bonkgalter,B., Brown,A., Burkett,G., Castle,A.,
              Chepel,V., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
              DeArnell,K., Dewar,K., Domino,M., Doyle,M., Fennestor,U.,
              Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,D.,
              Gaidyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
              Howland,T.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
              Landers,T., Lechoczky,J., Levine,R., Liu,C., Luo,G., Locke,K.,
              MacDonald,P., Marquis,N., McEwan,P., McCurt,A., McKernan,K.,
              McPherson,R., Meldrum,J., Menus,L., Morrow,D., Naylor,J.,
              Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
              Piatre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
              Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
              Stojanovic,N., Subramanian,A., Talamas,J., Teeffaye,S., Theodore,J.,
              Tirelli,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
              Zimmer,A. and Zody,W.
TITLE         Direct Submission
COMMENT       Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              On Jul 13, 2000 this sequence version replaced gi:6910620.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              -----
              Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              -----
              Project Information
              Center project name: L5106
              Center clone name: 511_E_14
              -----
              * NOTE: This record contains 86 individual
              * sequencing reads that have not been assembled into
              * contigs. Runs of N are used to separate the reads
              * and the order in which they appear is completely
              * arbitrary. Low-pass sequence sampling is useful for
              * identifying clones that may be gene-rich and allows
              * overlap relationships among clones to be deduced.
              * However, it should not be assumed that this clone
              * will be sequenced to completion. In the event that
              * the record is updated, the accession number will
              * be preserved.
              *
              1
              * 933: contig of 933 bp in length
              * 934 1033: gap of 100 bp
              * 1034 1961: contig of 928 bp in length
              * 1962 2061: gap of 100 bp
              * 2062 3005: contig of 944 bp in length
              * 3006 3105: gap of 100 bp
              * 3106 4066: contig of 961 bp in length
              * 4067 5126: gap of 100 bp
              * 5126: contig of 954 bp in length

```

5121	5120:	gap of 100 bp
5221	6142:	contig of 922 bp in length
6143	6242:	gap of 100 bp
6243	7207:	contig of 965 bp in length
7208	7307:	gap of 100 bp
7308	8234:	contig of 927 bp in length
8235	8334:	gap of 100 bp
8335	9276:	contig of 942 bp in length
9277	9376:	gap of 100 bp
9377	10329:	contig of 953 bp in length
10330	10429:	gap of 100 bp
10430	11370:	contig of 941 bp in length
11371	11470:	gap of 100 bp
11471	12433:	contig of 969 bp in length
12440	12539:	gap of 100 bp
12540	13474:	contig of 935 bp in length
13475	13574:	gap of 100 bp
13575	14538:	contig of 964 bp in length
14539	14638:	gap of 100 bp
14639	15599:	contig of 957 bp in length
15599	15695:	gap of 100 bp
15696	16621:	contig of 928 bp in length
16624	16723:	gap of 100 bp
16724	17659:	contig of 936 bp in length
17660	17759:	gap of 100 bp
17760	18702:	contig of 943 bp in length
18703	18802:	gap of 100 bp
18803	19857:	contig of 955 bp in length
19758	19957:	gap of 100 bp
19958	20787:	contig of 930 bp in length
20788	20887:	gap of 100 bp
20888	21836:	contig of 949 bp in length
21837	21936:	gap of 100 bp
21937	22865:	contig of 929 bp in length
22866	22965:	gap of 100 bp
22966	23906:	contig of 941 bp in length
23907	24006:	gap of 100 bp
24007	24929:	contig of 923 bp in length
24930	25029:	gap of 100 bp
25030	25954:	contig of 924 bp in length
25954	26054:	gap of 100 bp
26054	26992:	contig of 935 bp in length
26993	27092:	gap of 100 bp
27093	28049:	contig of 957 bp in length
28050	28149:	gap of 100 bp
28150	29108:	contig of 959 bp in length
29109	29208:	gap of 100 bp
29209	30124:	contig of 916 bp in length
30125	30224:	gap of 100 bp
30225	31124:	contig of 900 bp in length
31125	31224:	gap of 100 bp
31225	32179:	contig of 955 bp in length
32180	32279:	gap of 100 bp
32280	33207:	contig of 928 bp in length
33208	33307:	gap of 100 bp
33308	34249:	contig of 942 bp in length
34250	34349:	gap of 100 bp
34350	35289:	contig of 940 bp in length
35290	35389:	gap of 100 bp
35390	36327:	contig of 938 bp in length
36328	36427:	gap of 100 bp
36428	37353:	contig of 926 bp in length
37354	37453:	gap of 100 bp
37454	38408:	contig of 955 bp in length
38409	38508:	gap of 100 bp
38509	39427:	contig of 919 bp in length
39428	39527:	gap of 100 bp
39528	40459:	contig of 932 bp in length
40460	40559:	gap of 100 bp
40560	41483:	contig of 924 bp in length
41484	41589:	gap of 100 bp
41584	42543:	contig of 960 bp in length
42543	42643:	gap of 100 bp

[illegible]

[illegible]

[illegible]

Submitted (23-Oct-2003) Balzerque S., UMRV, INRA/CNRS, 2 rue  
directeur.sadmarc@inra.fr  
Gaston Crepeux, 91057 Evry cedex, FRANCE  
Gaston Crepeux on DNA from transformants of *Arabidopsis thaliana*  
plants from INRA (Versailles). The DNA fragment(s) resulting from  
the PCR were directly sequenced from the left or the right border  
to determine the genomic sequence flanking the insertion. T-DNA  
derived sequences were removed. Information to order the  
corresponding mutant line and a link to a database providing a  
graphical display of the insertion site are available at  
<http://dbsgap.versailles.inra.fr/publi/lines/>. This sequence has  
been generated in the framework of the French plant genomics  
program 'Genopiantes' (<http://www.genopiantes.com> and



http://genoplante-info.infobiogen.fr.

FEATURES  
Location/Qualifiers

source  
1..810  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Massiliaewskija"  
/db\_xref="taxon:3702"  
/clone="606H03"  
/clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
misc\_feature  
1..810  
/note="T-DNA flanking sequence  
right border"

## ORIGIN

Query Match 9.8%; Score 68.4; DB 8; Length 810;  
Best Local Similarity 41.2%; Pred. No. 0.00069;  
Matches 213; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

QY 175 TTCTCATTTGTCCTCTTCGTCAGTTATTTCTTCCCGAGTCCGACTCAGTAC 234  
DB 142 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 201  
QY 235 TCTCACTCTCGCGCGCTTAACTAGCTCCGCTGTTACTGTAAAGTTCTGCC 294  
DB 202 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 261  
QY 235 TTAGAGCCCTCCGATGCGCCACCGCATGATCTGCTCGATTTCTTTCTGCGT 354  
DB 262 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 321  
QY 355 GGAAAAATGACCTAATGTTCTCGATTGGAAGTTTGTGCTATGSGTTACTTTTC 414  
DB 322 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 381  
QY 415 CCTAATTTTATAGTTCTAGGTAACGATACCTGCGCTTACTGTTTGTTCATTGT 474  
DB 382 NTTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 441  
QY 475 TGTGCTTACGCGTTAGCGCTGATCGAGATTGACTGTGAAAAATCCTTCGTTT 534  
DB 442 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 501  
QY 535 TGGTTTGTGTTCAATATAAATGATGATACCTTTGCGCTTGAATGTTGTTT 594  
DB 502 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 561  
QY 595 GAGCTATGCGTGTGCTGTATTAACCTTCAAGTCAAGTGTGAGATTG 654  
DB 562 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 621  
QY 655 GTAGTACTGTGGGTTCTTGTGCTATAGTTGT 691  
DB 622 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 658

Search completed: March 29, 2004, 23:22:02  
Job time : 2626.78 secs

**This Page Blank (uspto)**







RESULT 5  
 US-10-424-599-73594/C  
 ; Sequence 73594, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 73594  
 ; LENGTH: 2890  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_37470C.1  
 US-10-424-599-73594

Query Match 6.0%; Score 197.8; DB 12; Length 2890;  
 Best Local Similarity 69.4%; Pred. No. 3.6e-39;  
 Matches 284; Conservative 0; Mismatches 122; Indels 3; Gaps 1;  
 QY 1243 GTGATGATTATGACTGCGATGTGATCAAAAGACCTGATCAGAAAGCAGATTAAGT 1302  
 DB 2890 GTGATGATTATGACTGCGATGTGATCAAAAGACCTGATCAGAAAGCAGATTAAGT 2831  
 QY 1303 GGTCAAAAAGTCTTTGGCAGCTTGTGATGCTGTGATCGAGCAGATTAAGTAAACAC 1362  
 DB 2830 GGTTAAGGATTTCTTTGAGAACTTGGATGCTGTGATCGAGCAGATTAAGTAAACAC 2771  
 QY 1363 AGAGCAGTGGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAAGTAAAC 1419  
 DB 2770 AAAGCAGTGGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAAGTAAAC 2711  
 QY 1420 ACCTGACCTCTACTAGCTCATGCGAGACAAAGAGCTAGCCGATTAAGTAAAC 1479  
 DB 2710 GACTGACGCTCTACTAGCTCATGCGAGACAAAGAGCTAGCCGATTAAGTAAAC 2651  
 QY 1480 GAGATTGGCTGAGATTGTAAGAAAGATCTACAGATGAGAGGCGATCTGCAATTCCT 1539  
 DB 2650 GGAAGTTGCTGAGCTTTTGAAGAGAACTGCGCAAAAGGCGCTTGCAGTAATTCAC 2591  
 QY 1540 GTGATGATTATGAGCAGTGAAGGCTTGGTGAAGATGAAGAAAGATTAAGTAAAC 1599  
 DB 2590 CTGGGGAAGTATTTGTAAGTGAAGGCTTTAAAGATGATGAAGAAAGATTAAGTAAAC 2531  
 QY 1600 TCTGGCCTCCAAAGTGCATCATGTAATCTAGTCACTGATTAAGAGACA 1648  
 DB 2530 TTTGGCCTCCAAAGTGCATCATGTAATCTAGTCACTGATTAAGAGAGACA 2482

RESULT 6  
 US-10-424-599-73591/C  
 ; Sequence 73591, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 73591  
 ; LENGTH: 1375

TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(1375)  
 ; OTHER INFORMATION: unsure at all n locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_37468C.1  
 US-10-424-599-73591

Query Match 5.2%; Score 169.8; DB 12; Length 1375;  
 Best Local Similarity 69.0%; Pred. No. 3e-32;  
 Matches 247; Conservative 0; Mismatches 108; Indels 3; Gaps 1;  
 QY 1243 GTGATGATTATGACTGCGATGTGATCAAAAGACCTGATCAGAAAGCAGATTAAGT 1302  
 DB 1375 GTGATGATTATGACTGCGATGTGATCAAAAGACCTGATCAGAAAGCAGATTAAGT 1316  
 QY 1303 GGTCAAAAAGTCTTTGGCAGCTTGTGATGCTGTGATCGAGCAGATTAAGTAAACAC 1362  
 DB 1315 GGTTAAGGATTTCTTTGAGAACTTGGATGCTGTGATCGAGCAGATTAAGTAAACAC 1256  
 QY 1363 AGAGCAGTGGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAAGTAAAC 1419  
 DB 1255 AAAGCAGTGGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAAGTAAAC 1196  
 QY 1420 ACCTGACCTCTACTAGCTCATGCGAGACAAAGAGCTAGCCGATTAAGTAAAC 1479  
 DB 1195 GACTGACGCTCTACTAGCTCATGCGAGACAAAGAGCTAGCCGATTAAGTAAAC 1136  
 QY 1480 GAGATTGGCTGAGATTGTAAGAAAGATCTACAGATGAGAGGCGATCTGCAATTCCT 1539  
 DB 1135 GGAAGTTGCTGAGCTTTTGAAGAGAACTGCGCAAAAGGCGCTTGCAGTAATTCAC 1076  
 QY 1540 GTGATGATTATGAGCAGTGAAGGCTTGGTGAAGATGAAGAAAGATTAAGTAAAC 1597  
 DB 1075 CTGGGGAAGTATTTGTAAGTGAAGGCTTTAAAGATGATGAAGAAAGATTAAGTAAAC 1018

RESULT 7  
 US-10-424-599-36215/C  
 ; Sequence 36215, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 36215  
 ; LENGTH: 1103  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(1103)  
 ; OTHER INFORMATION: unsure at all n locations  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_132704C.1  
 US-10-424-599-36215

Query Match 3.2%; Score 103.8; DB 12; Length 1103;  
 Best Local Similarity 67.3%; Pred. No. 1.7e-15;  
 Matches 175; Conservative 0; Mismatches 83; Indels 2; Gaps 2;  
 QY 2126 ATTGATTAGCAAAACAGGCTGAATTGAGTTGAATCATCAAGAG-ATGTTGTA 2184  
 DB 359 ATTCCCAAGTAAATCTAGACTCAAAATATGACATTAAGATCATCAAGAGATGTTGTA 300





PRIOR APPLICATION NUMBER: PCT/EP01/07537  
 PRIOR FILING DATE: 2001-07-02  
 PRIOR APPLICATION NUMBER: DE 10032529.7  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: DE 10043826.1  
 PRIOR FILING DATE: 2000-09-01  
 NUMBER OF SEQ ID NOS: 2424  
 SEQ ID NO 1670  
 LENGTH: 6668  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 FEATURE:  
 NAME/KEY: unsure  
 LOCATION: 1936  
 OTHER INFORMATION: n i s a o r g o r t  
 US-10-311-455-1670

Query Match 2.1%; Score 68.8; DB 14; Length 6668;  
 Best Local Similarity 45.8%; Pred. No. 5e-06;  
 Matches 238; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

QY 156 GTTTAGAGCTTAATAGCTCTCATTTGCTCTCTGCTGAGTTATTTCTCTC 215  
 |||||  
 DB 2742 GTTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2801  
 |||||  
 QY 216 CGAGCTCCTGACTACTACTCTGCTCCGCGCTTAACCTACGCTCCGCTT 275  
 |||||  
 DB 2802 TTTTGTGTTTTGTTTTTTTTTTTTTTTTTTTGTGTTTTTTTTTTT 2861  
 |||||  
 QY 276 ACTCTGAGTTTCTGCTTAGAGCTCCGATCCGCTCACCGAGCATTTGCTCG 335  
 |||||  
 DB 2862 TTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2921  
 |||||  
 QY 336 ATTTCTCTTTTCTGCTGAAAAATGCCCTAATGTTTCGATTCGAAAGTTTGT 395  
 |||||  
 DB 2922 TTTTGTGTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTT 2981  
 |||||  
 QY 396 GCTAGGTTACTTTTCCCTAATTTTATAGTTCTTAGTAACATACCTGCTCTTA 455  
 |||||  
 DB 2982 TTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3041  
 |||||  
 QY 456 CTTGTTTGTGCTATTTGTTGCTTTCACCGTTTAGTCGATCGAGATTTGACTG 515  
 |||||  
 DB 3042 TTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3101  
 |||||  
 QY 516 TGAATAATCCTGCTTTTGTGTTTGTTCATATAAATGAGATTGATCTACTTTGT 575  
 |||||  
 DB 3102 TGTGTTT 3161  
 |||||  
 QY 576 GCTTGAATGTTGTTTGTAGGCTAATGCTGTTGGCTGTATTAATCTACGTTGATG 635  
 |||||  
 DB 3162 TTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3221  
 |||||  
 QY 636 TGTGATTTTGAATTTGTGATGACTGTGGTTCTT 675  
 |||||  
 DB 3222 TTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3261  
 |||||

RESULT 11  
 US-10-239-676-52  
 Sequence 52, Application US/10239676  
 Publication No. US20030082609A1  
 GENERAL INFORMATION:  
 APPLICANT: OLEK, Alexander  
 APPLICANT: PIERENBROCK, Christian  
 APPLICANT: BERLIN, Kurt  
 TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
 FILE REFERENCE: 5013.1003  
 CURRENT APPLICATION NUMBER: US/10/239. 676  
 PRIOR FILING DATE: 2002-09-24  
 PRIOR APPLICATION NUMBER: PCT/EP01/03968

DE 10019058.8  
 DE 10019173.8  
 DE 10032529.7  
 DE 10043826.1  
 PRIOR FILING DATE: 2001-04-06  
 2000-04-06  
 2000-04-07  
 2000-06-30  
 2000-09-01  
 NUMBER OF SEQ ID NOS: 228  
 SEQ ID NO 52  
 LENGTH: 9539  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 FEATURE:  
 US-10-239-676-52

Query Match 2.0%; Score 66.8; DB 14; Length 9539;  
 Best Local Similarity 46.7%; Pred. No. 2e-05;  
 Matches 212; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

QY 251 TTTAACTTACGTTCTCCGTTGTTACTCTGTAAGTTTCTGCCTTAGAGCTCCGATCG 310  
 |||||  
 DB 202 TTTTGTGTTTTGTAATTTTGTATTTTGTGTTTTTTTTTTTATTTTCTTTC 261  
 |||||  
 QY 311 CCTCACCGATGCAATCTGCTCGATTTCTCTTTTCTGCTGAAAAATGCCCTAA 370  
 |||||  
 DB 262 GTTTTGTGTTTTGCTGTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTATTT 321  
 |||||  
 QY 371 TGTTCGATTTGGAAGTTTTTGCTATGAGTTACTTTTCCCTAATTTATAGTT 430  
 |||||  
 DB 322 TTAATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 381  
 |||||  
 QY 431 CTAGGAACGATACCTGCTCTTACTGTTTGTTCATTTGTTGCTTTCACCGTT 490  
 |||||  
 DB 382 TTTTGTGTTTTTTTTGTTTTTTTTTTTTTTTTTTTATTTTGTGTTTTTTTTT 441  
 |||||  
 QY 491 AGTCGCTGATCGAGATTTGACTGTGAAAAATCCTGTTTTGTTGTTTGTTCATA 550  
 |||||  
 DB 442 ATTTTGTGTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 501  
 |||||  
 QY 551 TAAATCGATGATCACTTTTGTGCTTTGATGTTTGTGTTGAGCTATGCTGTT 610  
 |||||  
 DB 502 TTTTGTGTTTTGATTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 561  
 |||||  
 QY 611 GCGTGTATACCTGATCGTTGATGTTGATTTGAGATTTGATGACTGAGGTT 670  
 |||||  
 DB 562 TTTTGTGTTTTGTTATTTTATTTTATTTTATTTTATTTTATTTATTTATTTT 621  
 |||||  
 QY 671 TCTTGGTGCTATAGGTTGTAATAATGAGTTCT 704  
 |||||  
 DB 622 TTAATTAATATTTTATTTTGTGAGTGAAGTTT 655  
 |||||

RESULT 12  
 US-10-240-453-54  
 Sequence 54, Application US/10240453  
 Publication No. US20030148326A1  
 GENERAL INFORMATION:  
 APPLICANT: OLEK, Alexander  
 APPLICANT: PIERENBROCK, Christian  
 APPLICANT: BERLIN, Kurt  
 TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
 TITLE OF INVENTION: Transcription  
 TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associate  
 TITLE OF INVENTION: with DNA Transcription  
 FILE REFERENCE: 5013.1009  
 CURRENT APPLICATION NUMBER: US/10/240.453  
 PRIOR FILING DATE: 2002-10-02  
 CURRENT APPLICATION NUMBER: PCT/EP01/03973  
 PRIOR FILING DATE: 2001-04-06  
 PRIOR APPLICATION NUMBER: DE 10019058.8

;; PRIOR FILING DATE: 2000-04-06  
;; PRIOR APPLICATION NUMBER: DE 10019173.8  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: DE 10032529.7  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: DE 10043826.1  
;; PRIOR FILING DATE: 2000-09-01  
;; NUMBER OF SEQ ID NOS: 350  
;; SEQ ID NO 54  
;; LENGTH: 9539  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-453-54

Query Match 2.0%; Score 66.8; DB 14; Length 9539;  
Best Local Similarity 46.7%; Pred. No. 2e-05;  
Matches 212; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

QY 251 TTTAACTTACGTTCTGCGGTTTACCTGAGTTTTCGCTTAGAGCCGCGATCG 310  
DB 202 TTTTCTTTTGTATTTTGTGTTTATTTTGTGTTTATTTTATTTTATTTTTC 261  
QY 311 CCTCACCAGTCATCTGCTGATTTCTTTTCTGCTGAAATTCCTTAA 370  
DB 262 GTTTTCTTTTGTGTTTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTT 321  
QY 371 TGTTCGATTTGCAAGTTTGTGCTTATGAGTTTCTTTTCCATATTTTATAGT 430  
DB 322 TTTATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 381  
QY 431 CTTAGAGTAAGATACCTGCTTACGTTTCTTCTTCTTCTTCTTCTTCTT 490  
DB 382 TTTTCTTTTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 441  
QY 491 AGTCGATCGAGATTTGACTGTGAAATCTTCTTCTTCTTCTTCTTCTTCT 550  
DB 442 ATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 501  
QY 551 TAAATCGATGATCTACCTTTGTGCTTGTGATTTGTTTGTGACCTTATGCT 610  
DB 502 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 561  
QY 611 GCGTGTATACCTGCTGATGCTGATTTTGTGATTTGTGATGCTGCTGCT 670  
DB 562 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 621  
QY 671 TCTTGTGCTATAGTTGTTAAATGAGTTCT 704  
DB 622 TTTATTTATATATTTTCTTTTGTAGGTTT 655

RESULT 13  
US-10-424-599-57885  
;; Sequence 57885, Application US/10424599  
;; Publication No. US20040031072A1  
;; GENERAL INFORMATION:  
;; APPLICANT: La Rosa Thomas J  
;; APPLICANT: Kovalic David K  
;; APPLICANT: Zhou Yihua  
;; APPLICANT: Cao Yongwei  
;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
;; FILE REFERENCE: 38-21(53223)B  
;; CURRENT APPLICATION NUMBER: US/10/424,599  
;; NUMBER OF SEQ ID NOS: 285684  
;; SEQ ID NO 57885  
;; LENGTH: 673  
;; TYPE: DNA  
;; ORGANISM: Glycine max  
;; FEATURE:

;; NAME/KEY: unsure  
;; LOCATION: (1)..(673)  
;; OTHER INFORMATION: unsure at all n locations  
;; FEATURE:  
;; OTHER INFORMATION: clone ID: PAT\_MRT3847\_23282C.1  
US-10-424-599-57885

Query Match 2.0%; Score 66.2; DB 12; Length 673;  
Best Local Similarity 47.0%; Pred. No. 4.8e-06;  
Matches 170; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 330 TGTGATTTCTCTTTTCTGCTGAAATTCCTATGTTCTGATTTGAAAGT 389  
DB 180 TGCACATCCCATTTACTGTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNN 239  
QY 390 TTTTGTCTATGAGTTTCTTTTCTTCTTATTTTATGTTCTTAGTACGATCTGC 449  
DB 240 TTTTCTTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 299  
QY 450 GTCTTACTGTTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 509  
DB 300 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 359  
QY 510 TGAATGAAATTCCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 569  
DB 360 TTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 419  
QY 570 TTTTGTCTTGTATGTTTGTGTTTGTGCTTATGCTTGTGCTTGTATTTACT 629  
DB 420 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 479  
QY 630 TTTATGATGATTTGATTTTGTGATGAGCTGAGCTGCTTCTTGTGCTATAGT 689  
DB 480 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 539  
QY 690 GT 691  
DB 540 AT 541

RESULT 14  
US-10-311-455-1931  
;; Sequence 1931, Application US/10311455  
;; Publication No. US20030143606A1  
;; GENERAL INFORMATION:  
;; APPLICANT: OLEK, Alexander  
;; APPLICANT: PIEPENBROCK, Christian  
;; APPLICANT: BERLIN, Kurt  
;; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter.  
;; TITLE OF INVENTION: Cytosine methylation  
;; FILE REFERENCE: 5013.1014  
;; CURRENT APPLICATION NUMBER: US/10/311,455  
;; CURRENT FILING DATE: 2002-12-16  
;; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
;; PRIOR FILING DATE: 2001-07-02  
;; PRIOR APPLICATION NUMBER: DE 10032529.7  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: DE 10043826.1  
;; PRIOR FILING DATE: 2000-09-01  
;; NUMBER OF SEQ ID NOS: 2424  
;; SEQ ID NO 1931  
;; LENGTH: 14006  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
;; NAME/KEY: unsure  
;; LOCATION: 8289, 8310, 8313  
;; OTHER INFORMATION: n 18 a or g or c or t  
US-10-311-455-1931

Query Match 2.0%; Score 66; DB 14; Length 14006;

Best Local Similarity 45.3%; Pred. No. 4.2e-05; Indels 0; Gaps 0;  
Matches 240; Conservative 0; Mismatches 290;

```

QY 157 TTTTAAAGCTAATAAGCTTCCATTTGCTCTTTCGTCAGTTATTTCTCTCC 216
    |||
Db 1974 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2033
QY 217 GAGTCCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 276
    |||
Db 2034 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2093
QY 277 CTCGTAAGTTTCTGCTTACGACGACGACGACGACGACGACGACGACGACG 336
    |||
Db 2094 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2153
QY 337 TTTTCCTTTTCTCGCGGAAATGACCCTAATGTCGATTCGAAAGTTTGTG 396
    |||
Db 2154 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2213
QY 397 CTATGCTTACTTTTCTCCATATTTTATTTTATTTTATTTTATTTTATTTT 456
    |||
Db 2214 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2273
QY 457 TGTTTTGTGATTTTGTGCTTTCACGCTTATGCTGATCGAGATTGACTGT 516
    |||
Db 2274 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2333
QY 517 GAAAAATCCTGCTTTTGTGCTTTCATATTAATGCAATGATTCACCTTTGTG 576
    |||
Db 2334 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2393
QY 577 CTTTATGTTTGTGTTTGGAGCTTATGCTTGTGCTTATTAATCTTCACTT 636
    |||
Db 2394 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2453
QY 637 GTGATTTTGAATTTGTGATGACGTGCTTCTTGTGCTATAG 686
    |||
Db 2454 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2503

```

RESULT 15  
US-10-311-455-2147  
; Sequence 2147, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPERBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 2147  
; LENGTH: 113515  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-2147

Query Match 2.0%; Score 65.8; DB 14; Length 113515;  
Best Local Similarity 48.5%; Pred. No. 0.0002; Indels 0; Gaps 0;  
Matches 181; Conservative 0; Mismatches 192;

QY 325 TTCTGTGCTGATTTCTTTCTTCGCTGAAATGCGCTAATGTTCTGATTTTCG 384

```

Db 51453 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 51512
QY 385 AAGTTTGTGCTAATGCTTATTTTCCATATTTTATTTTATTTTATTTTATTTT 444
    |||
Db 51513 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 51572
QY 445 CTTGCTTACTGTTTGTGCTTATTTGTTGCTTACGCTTATGCTGATCGGA 504
    |||
Db 51573 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 51632
QY 505 GTATTTGACGTGAAATCCCTTGTGCTTATTTGCTTATTAATGCAATTTGAT 564
    |||
Db 51633 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 51692
QY 565 CTACCTTTGTGCTTATGTTTGTGTTTATGACCTAATGCTTGTGCTTATTA 624
    |||
Db 51693 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 51752
QY 625 TCAGCTTCAATGCTGATTTTGAATTTGCTTATGCTGCTGCTTATTTGCTG 684
    |||
Db 51753 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 51812
QY 685 AGTTGTAAAT 697
    |||
Db 51813 AAGTTATAGAGT 51825

```

Search completed: March 30, 2004, 08:57:32  
Job time : 1121.16 secs



Best Local Similarity 6.4%; Pred. No. 1,6e-12;  
Matches 25; Conservative 232; Mismatches 131; Indels 0; Gaps 0;

[illegible]

RESULT 2  
US-09-313-294A-121

```

Patent No.6476212
GENERAL INFORMATION:
APPLICANT: Lalgudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 121
LENGTH: 272
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6476212 700548569H1
NAME/KEY: unsure
LOCATION: 14, 32, 61, 127, 162, 192
OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-121

```

Query Match 2.3%; Score 74.4; DB 4; Length 272;  
Best Local Similarity 58.7%; Pred. No. 2.4e-10;  
Matches 142; Conservative 0; Mismatches 97; Indels 3; Gaps 1;

QY	1777	TATGAGGCTCTTAGAGCAAGCGCATTTCCATATGTGCACAGGGGCANTCGTGGATATGATGTTT	1833
Db	2	TATGAAACCAAGTAAAGCACGTCATGCTCATATGCTCTTCCTGCGGCACCGGTGATATGAGCGSTN	61
QY	1837	CTGATGTTTGTGAGAGCAGTGCACACTGGCATTTTGGAGGCGCGAAGCGCTCCACCGGGAGTTA	1896
Db	62	TTATATATTTTGAAGCTCAGCTGTGGGGCTACATGGAAGTGAACGCTTCGATTAACACTTT	121
QY	1897	GCTGAGATGGGGGTAGATAGAAATTCCTCG---GGTCGAAGCGCAGATGTTTCTGGA	1953
Db	122	GTTTANCAAGTACAGACAGAAATTCATGCACTTCAGCAGGATTTGCAATTTGTGCTTGTT	181

[illegible]

RESULT 3  
US-09-313-294A-463

Patent No.6476212  
GENERAL INFORMATION:  
APPLICANT: Laigudi, Raghunath V.  
APPLICANT: Ito, Laura Y.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
NUMBER OF SEQ ID NOS: 7600  
SOFTWARE: PERL Program  
SEQ ID NO 463  
LENGTH: 274  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID NO. 6476212 700549278H1  
NAME/KEY: unsure  
LOCATION: 15, 53  
OTHER INFORMATION: a, t, c, g, or other  
US-09-313-294A-463

Query Match	2.2%	Score 72.2;	DB 4	Length 274;
Best Local Similarity	60.6%	Pred. No.9.6e-10;		
Matches 152; Conservative	0;	Mismatches 95;	Indels 4;	Gaps 2;

QY	1777	TATAGGCTCTTGAAGACGCAATTCTTAGT--CCAAGGGCCATGTGGAGTAGAGT	1835
Db	2	TATAGCAAGTANAGACAGTCATGCTTAGTCCCTCTGGGGCACCCTGGNATGAGCCT	61
QY	1836	TCTGATGTTTGAAGCAGTGCACCTGCTATTTTGAAGCCGACGCTTCAACGGAGTT	1895
Db	62	GTTAAATTTTGAAGCTCAGCTGGGGCTACATGAAGCTGACGTCTGCATAAACCTT	121
QY	1896	AGCTGAGATGGGGTTAGTATGAATTGGCTGG--GGTCAAGAGCGATATGTTTCTGG	1952
Db	122	TGTTATTCAGAGTACACACGAGATTATGGCACCTTACGCAAGTTTGATTTGTGCTTGG	181
QY	1953	AGGTGTTCCGCAACTGTATGGCTTCTTGCACAGACAGCAAGATCTGACATATTCAATCA	2012
Db	182	TGGGAAAGCAACTATATGTTCTTTAGCAAAACAAGAGGATATGAGGACATTAAACA	241
QY	2013	AGACTCTCAAG	2023
Db	242	GCATTGCCAGG	252

US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pc-Fls  
US-08-232-463-14

Query Match 2.1%; Score 68.6; DB 1; Length 7218;  
Best Local Similarity 9.0%; Pred. No. 6,1e-08;  
Matches 41; Conservative 230; Mismatches 184; Indels 0; Gaps 0;

2731 TTTCGACGCTGAGAGAGTGTCAAGCTTCATCGAGTTTCAAGAGAGAGATGAGAG 2790  
1442 TTTCGTCRRR 1363  
2791 TTTCGACGAGAGAGAGAGTGTCTAATAAGATCAAGAGAGAGATGAGAGATGAG 2850  
1382 RR 1323  
2851 AAGAGCATGACGAGAGATATTGATCTGAGAGAGAGATTTGATGAGGCTTGAACAG 2910  
1322 RRR 1263  
2911 CTCATGACAGAGATGCTTACACATGAGATGATGAGAGAGAGATGCTGACAGAG 2970  
1262 RRR 1203  
2971 ACAAGACTAAGTTCTTTGTTTGGTATGTCGAGAGAGAGATGAGAGATCTGAGAGAC 3030  
1202 RRR 1143  
3031 TCATTAATAATACAGAGACAATACTAAGAGATTAATATATCTCAATTTTAG 3090  
1142 RRR 1063  
3091 TAGACGATCTAAGAGAGATTAAGTTCTTGTGACTAAACCAAGTTCTCTAGATTTT 3150  
1082 RRR 1023  
3151 GTTTTTCGTAATAATTTCATATGAAAGTTAGA 3185  
1022 GTGAGCTATGCGACGAAGAGAGAAATAAGTTATA 988

RESULT 5  
US-09-331-581-3  
Sequence 3, Application US/09331581  
Patent No. 6130070  
GENERAL INFORMATION:  
APPLICANT: TOHDA, Hideki  
APPLICANT: HAMA, Yuko  
APPLICANT: KUMAGAI, Hiromichi  
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE  
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS  
TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF  
FILE REFERENCE: 0059-1142-0PCT  
CURRENT APPLICATION NUMBER: US/09/331,581  
CURRENT FILING DATE: 1999-06-30  
EARLIER APPLICATION NUMBER: PCT/JP98/04929  
EARLIER FILING DATE: 1998-10-30  
EARLIER APPLICATION NUMBER: JP 9-314608  
EARLIER FILING DATE: 1997-10-31  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 3  
LENGTH: 7286  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:DNA  
US-09-331-581-3

Query Match 1.6%; Score 52.6; DB 3; Length 7286;  
Best Local Similarity 56.7%; Pred. No. 0.0013;  
Matches 97; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

9 AACCAAAATTAGCAAGTCAATGTTCTGACATTAATTAATAGTGGAGCAATTAAGTTA 68  
4559 AACATACATGAGAGAAATGCTATGAGCTGAGAAATTAATGACGTCACAGCAAAAA 4618  
69 AGCGAAAAGAGAAAAAAGTACAAAATGAAAACAAATCAACTGAATGAATTT 128  
4619 AAG 4678  
129 GAGTCCAGATCGAGAAAAAGAGCGCGTTTATGAGCTTAATTAAGCTTCT 179  
4679 TACTTTGAGCGCGAGAAAGAACGCGGATCCAGACATGATTAAGATACAT 4729

RESULT 6  
US-09-331-581-14  
Sequence 14, Application US/09331581  
Patent No. 6130070  
GENERAL INFORMATION:  
APPLICANT: TOHDA, Hideki  
APPLICANT: HAMA, Yuko  
APPLICANT: KUMAGAI, Hiromichi  
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE  
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS  
TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF  
FILE REFERENCE: 0059-1142-0PCT  
CURRENT APPLICATION NUMBER: US/09/331,581  
CURRENT FILING DATE: 1999-06-30  
EARLIER APPLICATION NUMBER: PCT/JP98/04929  
EARLIER FILING DATE: 1998-10-30  
EARLIER APPLICATION NUMBER: JP 9-314608  
EARLIER FILING DATE: 1997-10-31  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 14  
LENGTH: 7938  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:DNA  
US-09-331-581-14

Query Match 1.5%; Score 52.6; DB 3; Length 7938;  
Best Local Similarity 56.7%; Pred. No. 0.0013;  
Matches 97; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 9 AACAAATTAAGCAAGTCAATGTTCTAGCAATTAATTAAGTGGAGCAATTAAGTGA 68  
DB 5211 AACATATCATGAGAAAGATGCTATGTAGTGAATAATGAATGATGACAGCAAAAA 5270  
QY 69 AGCGAAAAAGAAAAAGGTACAAAATGAAAACAAATCAACTGAATGAAAAATT 128  
DB 5271 AAG 5330  
QY 129 GAGATCGAGATCGAAAAACGAGCGGTTTAAAGCTTAATAGCTTCC 179  
DB 5331 TACCTTGTGAGGCGAAAGAACGACGAGATCCAGCATGATAGATACAT 5381

## RESULT 7

US-09-640-173-53  
Sequence 53, Application US/09640173  
Patent No. 6613515  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Stoik, John A.  
TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND  
FILE REFERENCE: 210121.484C2  
CURRENT APPLICATION NUMBER: US/09/640.173  
CURRENT FILING DATE: 2000-08-15  
NUMBER OF SEQ ID NOS: 196  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 53  
LENGTH: 396  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(396)  
OTHER INFORMATION: n = A,T,C or G  
US-09-640-173-53

Query Match 1.5%; Score 50.6; DB 4; Length 396;  
Best Local Similarity 49.8%; Pred. No. 0.00079;  
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 448  
DB 18 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 77  
QY 449 CGCTTACTGTTTGTTCATTTTGTGCTTACCGTTAGTCGTCATCGAGAT 508  
DB 78 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 137  
QY 509 TTGACTGAAAAATCCTGCTTTTGGTTTGTTCATATAATGATGATAC 568  
DB 138 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 197  
QY 569 CTTTGTGCTTGAAGTTGTTTGAAGCTTATGCGTTGCTGTTGTTAACTTCA 627  
DB 198 TTTTGTGCTTGAAGTTGTTTGAAGCTTATGCGTTGCTGTTGTTAACTTCA 256

## RESULT 8

US-09-713-550-53  
Sequence 53, Application US/09713550  
Patent No. 6617109  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Stoik, John A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.484C4

CURRENT APPLICATION NUMBER: US/09/713,550  
CURRENT FILING DATE: 2000-11-14  
NUMBER OF SEQ ID NOS: 205  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 53  
LENGTH: 396  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(396)  
OTHER INFORMATION: n = A,T,C or G  
US-09-713-550-53

Query Match 1.5%; Score 50.6; DB 4; Length 396;  
Best Local Similarity 49.8%; Pred. No. 0.00079;  
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 448  
DB 18 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 77  
QY 449 CGCTTACTGTTTGTTCATTTTGTGCTTACCGTTAGTCGTCATCGAGAT 508  
DB 78 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 137  
QY 509 TTGACTGAAAAATCCTGCTTTTGGTTTGTTCATATAATGATGATGATAC 568  
DB 138 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTAGTAAGTACGATACCTG 197  
QY 569 CTTTGTGCTTGAAGTTGTTTGAAGCTTATGCGTTGCTGTTGTTAACTTCA 627  
DB 198 TTTTGTGCTTGAAGTTGTTTGAAGCTTATGCGTTGCTGTTGTTAACTTCA 256

## RESULT 9

US-10-204-708-4  
Sequence 4, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
TITLE OF INVENTION: by Assessing DNA Methylation  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204.708  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 4  
LENGTH: 10619  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-4

Query Match 1.5%; Score 50.6; DB 4; Length 10619;  
Best Local Similarity 49.8%; Pred. No. 0.0055;  
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 404 TTACTTTTCCCTAATTTATAGTCTTGAAGTACGATACCTGCTTACTGTTT 463  
DB 128 TTTTGTGCTAAGGTTACTTTTCCCTAATTTATAGTCTTGAAGTACGATACCTG 197







schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/618,166  
FILING DATE: 17-Jul-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mcmasters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 240052,419C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 209:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51259 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-618-166-209  
SEQUENCE DESCRIPTION: SEQ ID NO: 209:  
Query Match 1.5%; Score 48.6; DB 4; Length 51259;  
Best Local Similarity 52.2%; Pred. No. 0.048;  
Matches 108; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 2764 GAGTTTCAGAGAAAGATGAGAGAGTTTGGAGAGAGAGATGCTGATTAAGAT 2823  
DB 232 GAGGAG 173  
QY 2824 CAG 2883  
DB 172 GAGCAG 113  
QY 2884 AAGAGATTGATGAGCTTTGGAACAGCTCATGTATCAAGCATGAGCTTTCATGAAGAT 2943  
DB 112 AAGAGAGGAG 53  
QY 2944 GATTGAGACAAAGCTGTGTACCAAG 2970  
DB 52 GAGGAG 26  
RESULT 15  
US-09-014-969-14/c  
Sequence 14, Application US/09014969  
Patent No. 5965397  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/014,969  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2447 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-014-969-14

Query Match 1.5%; Score 47.8; DB 2; Length 2447;  
Best Local Similarity 47.2%; Pred. No. 0.013;  
Matches 142; Conservative 1; Mismatches 158; Indels 0; Gaps 0;  
QY 390 TTTTGTCTATGAGTACTTTTCCCTATATTTATGTTCTTAGGTAACGATACCTGC 449  
DB 2447 TTTTGTCTATGAGTACTTTTCCCTATATTTATGTTCTTAGGTAACGATACCTGC 2388  
QY 450 GTCTTACTGTTTGTTCATTTTGTGCTTACCGTTAGTCGTCGATGCGAGTATT 509  
DB 2387 TTTTGTCTATGAGTACTTTTCCCTATATTTATGTTCTTAGGTAACGATACCTGC 2328  
QY 510 TCAGTGAAGAAATCCGTGTTTGTGTTTGTTCATATTAATCGATATGATCACC 569  
DB 2327 TTTTGTCTATGAGTACTTTTCCCTATATTTATGTTCTTAGGTAACGATACCTGC 2268  
QY 570 TTTTGTCTATGAGTACTTTTGTGAGCTATGCGTTGCTGTTATTAATCTCACG 629  
DB 2267 TTTTGTCTATGAGTACTTTTGTGAGCTATGCGTTGCTGTTATTAATCTCACG 2208  
QY 630 TTGATGTGAGATTGAGATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 669  
DB 2207 ATCAATATACAGATCCAGATATGTGAACCATATATATATATATATATATATAT 2148  
QY 690 G 690  
DB 2147 G 2147

Search completed: March 30, 2004, 03:14:55  
Job time : 249.409 secs

**This Page Blank (uspto)**



XX Beclin C, Elmayer T, Vaucheret H;  
 XX WPI; 2001-159529/16.  
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 PT resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 XX  
 PS Claim 1; Page 31-32; 36pp; French.  
 XX  
 CC The present sequence represents the genomic sequence of the Arabidopsis  
 CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional  
 CC inactivation (degradation of RNA) and for resistance to viruses.  
 CC Overexpression of SGS3 results in plants with increased resistance to  
 CC viruses, while inactivation of SGS3 in transgenic plants (e.g. by  
 CC expressing antisense RNA, by mutation or by homologous recombination)  
 CC increases the level of the transgene product. This product may e.g.  
 CC impart resistance to herbicide, insects or pathogens, alter contents of  
 CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an  
 CC immunoglobulin or interferon  
 XX  
 S0 Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 3275; DB 4; Length 3275;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAACAAACAAATTTAGCAATGTCATGTTCTGACATTAATTAATAGTGGAAACA 60  
 DB 1 GACAAACAAACAAATTTAGCAATGTCATGTTCTGACATTAATTAATAGTGGAAACA 60  
 QY 61 TTAAGTTAAGCGAAAAAGAAAAAGTACAAAAATGAAAAATTAATCTAAT 120  
 DB 61 TTAAGTTAAGCGAAAAAGAAAAAGTACAAAAATGAAAAATTAATCTAAT 120  
 QY 121 GAAATTTGAGTCCAGATCGAAAAACGAGCGCTTTTAGACTTAATAGCTTCCTC 180  
 DB 121 GAAATTTGAGTCCAGATCGAAAAACGAGCGCTTTTAGACTTAATAGCTTCCTC 180  
 QY 121 GAAATTTGAGTCCAGATCGAAAAACGAGCGCTTTTAGACTTAATAGCTTCCTC 180  
 DB 121 GAAATTTGAGTCCAGATCGAAAAACGAGCGCTTTTAGACTTAATAGCTTCCTC 180  
 QY 181 ATTTGCTCTTCTGTCGATTAATTTCTCTCCGAGTCTGACTCACTCTCAC 240  
 DB 181 ATTTGCTCTTCTGTCGATTAATTTCTCTCCGAGTCTGACTCACTCTCAC 240  
 QY 241 TCTCCGCGCTTTAACTAGTCTCCGCTTAACTCTAGTATTTTCGCTTAGAG 300  
 DB 241 TCTCCGCGCTTTAACTAGTCTCCGCTTAACTCTAGTATTTTCGCTTAGAG 300  
 QY 241 TCTCCGCGCTTTAACTAGTCTCCGCTTAACTCTAGTATTTTCGCTTAGAG 300  
 DB 241 TCTCCGCGCTTTAACTAGTCTCCGCTTAACTCTAGTATTTTCGCTTAGAG 300  
 QY 301 CCTCCGATCGCCTCAACGATGCAATTCGTGCTGATTTCTTTTCTGCTGAAAA 360  
 DB 301 CCTCCGATCGCCTCAACGATGCAATTCGTGCTGATTTCTTTTCTGCTGAAAA 360  
 QY 361 ATTGCCCTAATGTTCTGATTTGCAAGTTTTTGTCTAGAGGTTACTTTTCCCTATA 420  
 DB 361 ATTGCCCTAATGTTCTGATTTGCAAGTTTTTGTCTAGAGGTTACTTTTCCCTATA 420  
 QY 421 TTTTATAGTCTTGGTAAAGATACCTGCGCTTACTAGTTTGTCTATTTTGTGCT 480  
 DB 421 TTTTATAGTCTTGGTAAAGATACCTGCGCTTACTAGTTTGTCTATTTTGTGCT 480  
 QY 481 TTCAACGTTTAGTGGCTGATCGAGTATTTGACTGTGAAAAATCCTTGTTTTGGTTT 540  
 DB 481 TTCAACGTTTAGTGGCTGATCGAGTATTTGACTGTGAAAAATCCTTGTTTTGGTTT 540  
 QY 541 TTGTTTCAATATAATCGATGATCTATTTGCTGCTTGAATGTTTGTGAGCCT 600  
 DB 541 TTGTTTCAATATAATCGATGATCTATTTGCTGCTTGAATGTTTGTGAGCCT 600  
 QY 601 ATGGGTTGTTGCTTTGTTAATACTCAAGTCAATGATGATGATTTGATGATG 660  
 DB 601 ATGGGTTGTTGCTTTGTTAATACTCAAGTCAATGATGATGATGATTTGATGATG 660  
 QY 661 ACTGTGGTTTCTTTGTTGCTATAGTTGTAATAATGATTTCTAGGCTGTCCAATGT 720

DB 661 ACTGTGGTTTCTTTGTTGCTATAGTTGTAATAATGATTTCTAGGCTGTCCAATGT 720  
 QY 721 CTAAAGAAAAAAGACCTTACAGGGTGTATTAAGGCTTGAACAGTTGTTCAAGTT 780  
 DB 721 CTAAAGAAAAAAGACCTTACAGGGTGTATTAAGGCTTGAACAGTTGTTCAAGTT 780  
 QY 781 TGGAGGAGACGAGACGCTTCTTCAACAGATGATGAGAGAGTGGAGGTCATTCCA 840  
 DB 781 TGGAGGAGACGAGACGCTTCTTCAACAGATGATGAGAGAGTGGAGGTCATTCCA 840  
 QY 841 AGAAGAACAAAGCAAAACGAGAACACTTCTGAAAAAATTGGGTTTCTCAGAAATTCGA 900  
 DB 841 AGAAGAACAAAGCAAAACGAGAACACTTCTGAAAAAATTGGGTTTCTCAGAAATTCGA 900  
 QY 901 ATCTCTTAGAGCTTGGGGTGTGTCACAGACGAAAGAGAGTACCAACGATCTGAGAG 960  
 DB 901 ATCTCTTAGAGCTTGGGGTGTGTCACAGACGAAAGAGAGTACCAACGATCTGAGAG 960  
 QY 961 GAAACAAATGATCCGGGAGAGTAAACGCAATGTCGGGGCATTTCAAGCTAATATCTG 1020  
 DB 961 GAAACAAATGATCCGGGAGAGTAAACGCAATGTCGGGGCATTTCAAGCTAATATCTG 1020  
 QY 1021 GTCCGGGACGAGCGTTGAGCAGAAAGTATGATTAACAATTGTGGAACCCCACTGTAT 1080  
 DB 1021 GTCCGGGACGAGCGTTGAGCAGAAAGTATGATTAACAATTGTGGAACCCCACTGTAT 1080  
 QY 1081 CTGGCCTCTCTTTGAAAGAGATGGAATTGGCAGCAGAGAGGTTCTGCTCAGACA 1140  
 DB 1081 CTGGCCTCTCTTTGAAAGAGATGGAATTGGCAGCAGAGAGGTTCTGCTCAGACA 1140  
 QY 1141 CAGCTGTGAGAGATTCTGAGCTGAGAGATGATGATGATGATGATGATGATGATGATG 1200  
 DB 1141 CAGCTGTGAGAGATTCTGAGCTGAGAGATGATGATGATGATGATGATGATGATGATG 1200  
 QY 1201 ATGATTCGATGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 DB 1201 ATGATTCGATGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 QY 1261 ATGTGAGTCAAAAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATG 1320  
 DB 1261 ATGTGAGTCAAAAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATG 1320  
 QY 1321 GCACTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 DB 1321 GCACTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 QY 1381 CAGCTTGTCAAGAACGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1440  
 DB 1381 CAGCTTGTCAAGAACGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1440  
 QY 1441 ATGCGAGAGCAAAAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
 DB 1441 ATGCGAGAGCAAAAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
 QY 1501 AAAAGGATCTACAGATGAGAGGCGCATCTGCAATTCCTTGTGAGGATTTAAGGAGT 1560  
 DB 1501 AAAAGGATCTACAGATGAGAGGCGCATCTGCAATTCCTTGTGAGGATTTAAGGAGT 1560  
 QY 1561 GAAAGGTTTGGTGAAGATGAAGAGATTAAGAAATGCTGAGCTTCAATGATGATCA 1620  
 DB 1561 GAAAGGTTTGGTGAAGATGAAGAGATTAAGAAATGCTGAGCTTCAATGATGATCA 1620  
 QY 1621 TCAATGATCTACAGATGAGATGAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 1680  
 DB 1621 TCAATGATCTACAGATGAGATGAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 1680  
 QY 1681 TTTCTTAATTTTCTGTCATTTCTAATGATGATGATGATGATGATGATGATGATGATG 1740  
 DB 1681 TTTCTTAATTTTCTGTCATTTCTAATGATGATGATGATGATGATGATGATGATGATG 1740  
 QY 1741 ATGGGCAACCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800

Db 1741 ATGGGCAACCAAGAGCTGCTGGAATCTTCACAAAGTATGAGGCTCTTAGACACGCGCAT 1800  
 Qy 1801 TCCATAGGTCCAGAGGGGCTATCGTGGGATGAGTGTTCGATGTTTGAAGAGCGCCACT 1860  
 Db 1801 TCCATAGGTCCAGAGGGGCTATCGTGGGATGAGTGTTCGATGTTTGAAGAGCGCCACT 1860  
 Qy 1861 GGCATATTTGAGGCGGACGCGCTCCACCGGAGTTAGCTGAGATGGGTTAGATGAAT 1920  
 Db 1861 GGCATATTTGAGGCGGACGCGCTCCACCGGAGTTAGCTGAGATGGGTTAGATGAAT 1920  
 Qy 1921 GCTGGGGGTCAAGAGCGCATGATGTTTCTGAGAGTGTTCGCAACTGTATGCTTCTT 1980  
 Db 1921 GCTGGGGGTCAAGAGCGCATGATGTTTCTGAGAGTGTTCGCAACTGTATGCTTCTT 1980  
 Qy 1981 GCAAGCAAGCAAGATCTGACATATTCATCAACCTCAAGTCTCTCCCAAGA 2040  
 Db 1981 GCAAGCAAGCAAGATCTGACATATTCATCAACCTCAAGTCTCTCCCAAGA 2040  
 Qy 2041 AATTGATATATGCTTTAGTTTGTCTATGGAATTTAAAGTTTGTGTCGGTGTAA 2100  
 Db 2041 AATTGATATATGCTTTAGTTTGTCTATGGAATTTAAAGTTTGTGTCGGTGTAA 2100  
 Qy 2101 TGCACTGTTTATGATATATCTATGATTCATTAGGCAAAACAAGGCTGAAATTCAGATTG 2160  
 Db 2101 TGCACTGTTTATGATATATCTATGATTCATTAGGCAAAACAAGGCTGAAATTCAGATTG 2160  
 Qy 2161 AAATCATACCAAGAGATGTTGTAAGAGCTGAGGCAATCTCTGAGGCAATTCAGAG 2220  
 Db 2161 AAATCATACCAAGAGATGTTGTAAGAGCTGAGGCAATCTCTGAGGCAATTCAGAG 2220  
 Qy 2221 CTGAACCTATTTAAGAACCAAGCTCTCAAAAAGAACCAAGCAGGCTCTTGAGGA 2280  
 Db 2221 CTGAACCTATTTAAGAACCAAGCTCTCAAAAAGAACCAAGCAGGCTCTTGAGGA 2280  
 Qy 2281 TCTCTGGAATTTATGAGGAGAGAGCTGCTGTAAGAACTGACAGATATCGATGCTGAGA 2340  
 Db 2281 TCTCTGGAATTTATGAGGAGAGAGCTGCTGTAAGAACTGACAGATATCGATGCTGAGA 2340  
 Qy 2341 CAGAGACTAAGATGACGACATGAACAGAACGGAAGGATATATTTCTTGAAAT 2400  
 Db 2341 CAGAGACTAAGATGACGACATGAACAGAACGGAAGGATATATTTCTTGAAAT 2400  
 Qy 2401 CACAACTTGACATTTTGTATTTTACCTACTGATTCACATTTTGTATTAATGTCACAA 2460  
 Db 2401 CACAACTTGACATTTTGTATTTTACCTACTGATTCACATTTTGTATTAATGTCACAA 2460  
 Qy 2461 AAAAAGCTGCTGCTTTGAAGATGATGACACAGACAGGTTTTCATGATTCATCAAA 2520  
 Db 2461 AAAAAGCTGCTGCTTTGAAGATGATGACACAGACAGGTTTTCATGATTCATCAAA 2520  
 Qy 2521 CAGATCATGAAAG 2580  
 Db 2521 CAGATCATGAAAG 2580  
 Qy 2581 CGTGCCAAAGGTTTGGCCACAGCAGACAGACAACTTAATCCCTCTAGCAATGAGATTGC 2640  
 Db 2581 CGTGCCAAAGGTTTGGCCACAGCAGACAGACAACTTAATCCCTCTAGCAATGAGATTGC 2640  
 Qy 2641 CGAAAGAGTATATGATCTAACTAACTAATATCCCTCTGCGCTTTTGTTCAAACTTA 2700  
 Db 2641 CGAAAGAGTATATGATCTAACTAACTAATATCCCTCTGCGCTTTTGTTCAAACTTA 2700  
 Qy 2701 AGAGTACTGAATTAATCCGCTTTTGAATCTTTTGGCAGAGCTGAGAGAGTCAAGCTTC 2760  
 Db 2701 AGAGTACTGAATTAATCCGCTTTTGAATCTTTTGGCAGAGCTGAGAGAGTCAAGCTTC 2760  
 Qy 2761 ATCGAGTTTCAAGAGAGAGAGATGAGAGAGTTTGTGAAGAGAGAGAGATGCTGATAAA 2820  
 Db 2761 ATCGAGTTTCAAGAGAGAGAGATGAGAGAGTTTGTGAAGAGAGAGAGATGCTGATAAA 2820  
 Qy 2821 GATCAAGAGAGAGAGATGAGAGAGATGAGAGAGAGAGAGAGATGAGAGATTTGATCTG 2880  
 Db 2821 GATCAAGAGAGAGAGATGAGAGAGATGAGAGAGAGAGAGAGATGAGAGATTTGATCTG 2880

Qy 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCAATGATACAGCATGAGGCTTCAAGTGA 2940  
 Db 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCAATGATACAGCATGAGGCTTCAAGTGA 2940  
 Qy 2941 GATGATTTGACAAAGAGCTTGTGATACAGACAGAGACTAATGTTCTTTGTTTGGTTTG 3000  
 Db 2941 GATGATTTGACAAAGAGCTTGTGATACAGACAGAGACTAATGTTCTTTGTTTGGTTTG 3000  
 Qy 3001 GTATGTGGAAAGTAGAGAGATCTGAGAGACTCCATTTAATATCTAGCAATCTAAGGA 3060  
 Db 3001 GTATGTGGAAAGTAGAGAGATCTGAGAGACTCCATTTAATATCTAGCAATCTAAGGA 3060  
 Qy 3061 GATTATGATATATATCTCCATTTTATGAGAGAGATCTAAGAGACTTAAGTCTT 3120  
 Db 3061 GATTATGATATATATCTCCATTTTATGAGAGAGATCTAAGAGACTTAAGTCTT 3120  
 Qy 3121 GTGACTAAGCAAGTTTCTTATGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 3180  
 Db 3121 GTGACTAAGCAAGTTTCTTATGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 3180  
 Qy 3181 TTAGACATTTTACCAAGCTGACAGTGAATGACAGATGCAATTCAAATCATGTTTTT 3240  
 Db 3181 TTAGACATTTTACCAAGCTGACAGTGAATGACAGATGCAATTCAAATCATGTTTTT 3240  
 Qy 3241 AGAATTTATATCTCAAAAATATATGTTGGTCAAT 3275  
 Db 3241 AGAATTTATATCTCAAAAATATATGTTGGTCAAT 3275  
 RESULT 2  
 ID AAF25374 standard; cDNA; 1878 BP.  
 XX AAF25374;  
 AC 15-MAY-2001 (first entry)  
 DT 15-MAY-2001 (first entry)  
 XX  
 DE Nucleotide sequence of the Arabidopsis SGS3 polypeptide.  
 XX  
 KW SGS3 gene; post-transcriptional inactivation; RNA degradation;  
 KW viral resistance; resistance; fatty acid content; protein content; ss.  
 OS Arabidopsis thaliana.  
 XX  
 FT Key  
 FT CDS Location/Qualifiers  
 FT 1..1878  
 FT /\*cag= a  
 FT /product= "SGS3"  
 XX  
 EN W0200105951-A2.  
 XX  
 PD 25-JAN-2001.  
 XX  
 PF 13-JUL-2000; 2000WO-FR002052.  
 XX  
 PR 16-JUL-1999; 99FR-00009417.  
 PR 26-JAN-2000; 2000FR-00001006.  
 XX  
 PA (AVET ) AVENTIS CROSCIENCE SA.  
 PA (INRG ) INST NAT RECH AGRONOMIQUE.  
 PA  
 PI Beclin C, Elmayer T, Vaucheret H;  
 XX  
 DR WPI; 2001-159529/16.  
 DR P-PSDB; AAB31798.  
 XX  
 PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 PT resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 XX  
 PS Claim 1; Page 32-35; 36pp; French.  
 XX



CC The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.  
 CC The SGS3 gene is essential for post-transcriptional inactivation  
 CC (degradation of RNA) and for resistance to viruses. Overexpression of  
 CC SGS3 results in plants with increased resistance to viruses, while  
 CC inactivation of SGS3 in transgenic plants (e.g. by expressing an antisense  
 CC RNA, by mutation or by homologous recombination) increases the level of  
 CC the transgene product. This product may e.g. impart resistance (to  
 CC herbicide, insects or pathogens), alter contents of essential fatty acids  
 CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or  
 CC interferon

XX Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 44.6%; Score 1462; DB 4; Length 1878;  
 Best Local Similarity 83.3%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 376; Gaps 4;

QY 696 ATGAGTTCTAGGGCTGCTCCATGTCTAAAGAAAGAAAGCTTCAGGGTGTATAGGCT 755  
 Db 1 ATGAGTTCTAGGGCTGCTCCATGTCTAAAGAAAGAAAGCTTCAGGGTGTATAGGCT 60  
 QY 756 GAGGTGAACAGTGGTCAAGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 815  
 Db 61 GAGGTGAACAGTGGTCAAGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 QY 816 GAGAGAGAGTGGAGAGTCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875  
 Db 121 GAGAGAGAGTGGAGAGTCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 QY 876 AAAACCTGGGTTTCTCAGAAATTCGAATCTCTAGAGCTTGGGGTGTGTAGAGAGAG 935  
 Db 181 AAAACCTGGGTTTCTCAGAAATTCGAATCTCTAGAGCTTGGGGTGTGTAGAGAGAG 240  
 QY 936 AGAGTAGAGACGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 995  
 Db 241 AGAGTAGAGACGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 QY 996 CGGGGCAATCAAGCTAATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1055  
 Db 301 CGGGGCAATCAAGCTAATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 QY 1056 AACCTTGGGAG 1115  
 Db 361 AACCTTGGGAG 420  
 QY 1116 GCAAGAGAGAGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1175  
 Db 421 GCAAGAGAGAGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 QY 1176 GTGATTAATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1235  
 Db 481 GTGATTAATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 QY 1236 CTTGCAAGTGAATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1295  
 Db 541 CTTGCAAGTGAATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 QY 1296 AATAAGTGTCAAAAAGTCTTTGGAGAGTGTGAATGCTTGTGATCGAGAGAGTAAT 1355  
 Db 601 AATAAGTGTCAAAAAGTCTTTGGAGAGTGTGAATGCTTGTGATCGAGAGAGTAAT 660  
 QY 1356 GAACACAG 1415  
 Db 661 GAACACAG 720  
 QY 1416 TATAAAGTGAACCTCTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1475  
 Db 721 TATAAAGTGAACCTCTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 QY 1476 CATAGAGAGAGTGGCTGAAGTTTGAAGAGAGATCTAAGATGAGAGAGAGAGAGAG 1535  
 Db 781 CATAGAGAGAGTGGCTGAAGTTTGAAGAGAGATCTAAGATGAGAGAGAGAGAGAGAG 840

QY 1536 CTTGTGTGAGATTTATGAG 1595  
 Db 841 CTTGTGTGAGATTTATGAG 900  
 QY 1596 ATTGTCTGGCTCCAAAGTGTATATATATATATATATATATATATATATATATATAT 1655  
 Db 901 ATTGTCTGGCTCCAAAGTGTATATATATATATATATATATATATATATATATATAT 960  
 QY 1656 AAGGTGAATCTTCTGTCTTTACTTCTTAATTTTCTTGTGACATTCATGATCTTA 1715  
 Db 961 A----- 961  
 QY 1716 GAATTTAATTTGATGCTCGCATGGCAACAAAGAGCTCTGAAATCTTGACAA 1775  
 Db 962 -----AGTGGCTCGCATGGCAACAAAGAGCTCTGAAATCTTGACAA 1007  
 QY 1776 GTATGAGAGCTCTTATGAG 1835  
 Db 1008 GTATGAGAGCTCTTATGAG 1067  
 QY 1836 TTTGATGTTTGAAG 1895  
 Db 1068 TTTGATGTTTGAAG 1127  
 QY 1896 AGCTGAGATGGGTTTGAATGAAATTTGCTGGGCTCAAGAGCCGATATGTTTCTGAGAG 1955  
 Db 1128 AGCTGAGATGGGTTTGAATGAAATTTGCTGGGCTCAAGAGCCGATATGTTTCTGAGAG 1187  
 QY 1956 TGTTCGCAACCTGATGAGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2015  
 Db 1188 TGTTCGCAACCTGATGAGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1247  
 QY 2016 CTCTCAAGGTTCTCTCCCAAGAAATTTGATATATGCTTTAGTTTGTATTTGAAT 2075  
 Db 1248 CTCTCA----- 1253  
 QY 2076 TTAAGTTTGTGTCCTGCTTAATGATCTGTTATGATATATATATATATATATATAG 2135  
 Db 1254 -----AGG 1256  
 QY 2136 CAAG 2195  
 Db 1257 CAAG 1316  
 QY 2196 GCAAGTCTCTGAG 2255  
 Db 1317 GCAAGTCTCTGAG 1376  
 QY 2256 CAAG 2315  
 Db 1377 CAAG 1436  
 QY 2316 TCGAGAGATTAATCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2375  
 Db 1437 TCGAGAGATTAATCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496  
 QY 2376 AGAGGATATATTTCTTCAAGAAATCAAGAACTTGTATTTACTTCACTGATTTCA 2435  
 Db 1497 AG----- 1498  
 QY 2436 CATTTTGAATTAATTTGCAACAAAAAAGTGTGTGAGAGAGAGAGAGAGAGAGAGAG 2495  
 Db 1499 -----AGATGATGACACAG 1514  
 QY 2496 CAGGTTTTTCAATGATTTCAATCAACAGATCAATGAAAGAGAGAGAGAGAGAGAGAG 2555  
 Db 1515 CAGGTTTTTCAATGATTTCAATCAACAGATCAATGAAAGAGAGAGAGAGAGAGAGAG 1574  
 QY 2556 TTTGAGATGTTTGCAG 2615  
 Db 1575 TTTGAGATGTTTGCAG 1674  
 QY 2616 TTAATCCTTCAAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2675

```

Db      1635 TAAATCCCTCTAGCAATGACGATGCGCAAG----- 1665
Qy      2676 CTGGCGTTTTTGTTCCTAACCTAAGAGTAACTGATTAATTCGGTTTGATTTCTTCG 2735
Db      1666 ----- 1665
Qy      2736 CAGAGCTGAGAGAGATGTCAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGAGTTTGT 2795
Db      1666 -AGAGCTGAGAGAGATGTCAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGAGTTTGT 1724
Qy      2796 GGAAGAGAGAGAGATGTCGATTAAGATCAAGAGAGAGATGAGAGAGATGAGAGAG 2855
Db      1725 GGAAGAGAGAGAGATGTCGATTAAGATCAAGAGAGAGATGAGAGAGATGAGAGAG 1784
Qy      2856 GCATCAGAGAGAGATGATTTGATCTGAGAGAAAGATTTGATGAGCTTTGAGAGAGCTCAT 2915
Db      1785 GCATCAGAGAGAGATGATTTGATCTGAGAGAAAGATTTGATGAGCTTTGAGAGAGCTCAT 1844
Qy      2916 GTACAGAGATGAGCTTCACATGAGATGATTTGA 2949
Db      1845 GTACAGAGATGAGCTTCACATGAGATGATTTGA 1878

RESULT 3
ABX81661
ID      ABX81661 standard; cDNA; 272 BP.
XX
AC      ABX81661;
XX
DT      24-APR-2003 (first entry)
XX
DE      Corn ear-derived polynucleotide (cpd) #121.
XX
KW      Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;
KW      structural gene; functional gene; regulatory gene;
KW      corn ear-specific profile; gene transcription; gene expression;
KW      hybrid plant; desirable trait expression; plant breeding program;
KW      inheritance; desired characteristic; growth; development; yield;
KW      disease resistance; environmental adaptability; quality; yield;
KW      multigene trait; plant; gene; ss.
XX
OS      Zea mays.
XX
PN      US6476212-B1.
XX
PD      05-NOV-2002.
XX
PF      14-MAY-1999; 99US-00313294.
XX
PR      26-MAY-1998; 98US-0086722P.
XX
PA      (INCY-) INCYTE GENOMICS INC.
XX
PI      Lalgudi RV, Ito LY, Sherman BK;
XX
DR      WPI; 2003-208840/20.
XX
PT      Novel purified corn-ear derived polynucleotide useful as hybridization
PT      probe for detecting polynucleotide in sample, and for identifying,
PT      evaluating, and altering desired characteristics associated with growth,
PT      development.
XX
PS      Example; SEQ ID NO 121, 390pp; English.
XX
CC      The present invention relates to the isolation of corn ear-derived
CC      polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022
CC      and SATMON023. Some of the cdps uniquely identify structural, functional,
CC      and regulatory genes of corn ear. The polynucleotide sequences are
CC      useful for detecting cdps in a sample, for producing a corn ear-specific
CC      profile of gene transcription, for detecting altered gene expression in
CC      hybrid or hybrid plants, and for screening several molecules for specific
CC      binding to the polynucleotide. The cdps are useful to identify, isolate,

```

```

CC      or extend identical or related corn-ear nucleic acid sequences from DNA
CC      libraries, and in nucleic acid amplification or hybridization techniques
CC      to follow the expression of desirable traits through plant breeding
CC      programs. Preferably, the cdps are used to identify, evaluate, alter, or
CC      follow the inheritance of desired characteristics associated with growth,
CC      and development, disease resistance, environmental adaptability, quality,
CC      and yield of corn. The cdps are also useful as molecular markers for
CC      studying inheritance and multigene traits in a plant breeding program.
CC      The cdps are useful for producing purified corn-ear polypeptides by
CC      recombinant techniques. They are also useful in diagnostic assays to
CC      detect or confirm conditions or diseases associated with abnormal levels
CC      of cdp expression. ABX81541-ABX89140 represent corn ear-derived
CC      polynucleotides (cdps) of the invention. Note: The sequence data for this
CC      patent did not form part of the printed specification, but was obtained
CC      in electronic format directly from the USPTO web site at
CC      seqdata.uspto.gov/peispidentry.html
XX
SQ      Sequence 272 BP; 79 A; 52 C; 68 G; 67 T; 0 U; 6 Other;

Query Match      2.3%; Score 74.4; DB 7; Length 272;
Best Local Similarity 58.7%; Pred. No. 7.8e-08;
Matches 142; Conservative 0; Mismatches 97; Indels 3; Gaps 1;

Qy      1777 TATGAGGCTCTTAGAGCAGCGCATTCCTATGCTCAGAGGCCATCGTGGATGAGTGT 1836
Db      2 TATGAGCAGATMAAGACGCGCATGCTATGCTCCTGCGGACCGGTGATGAGGCTN 61
Qy      1837 CTGATGTTTGAAGAGCAGTGCCTGCTATTTGAGAGCCGAAACGCTTCACCGGAGATTA 1896
Db      62 TTATATTTTAAAGCTGAGCTGTGGGCTTAACATGAACTGAACCTCTGCAATMAACACTTT 121
Qy      1897 GCTGAGATGGGGTTAGATAGATTTGCTGG---GCTGAGAGCGCAGATGATTTTCTGGA 1953
Db      122 GTTAAACAGGTACAGACAGGAATTCATGGCACTTACGANGTTCGATTTGCTCGGT 181
Qy      1954 GGTGTTGCGCACTGTATGCTTCCTTCAACGAGAGATCTGACATATTCATCA 2013
Db      182 GCGAAAGAGGCGCATATGATGTTCTTAAACAAAGAGATATGACGATTTAAACAG 241
Qy      2014 CA 2015
Db      242 CA 243

RESULT 4
ABX82003
ID      ABX82003 standard; cDNA; 274 BP.
XX
AC      ABX82003;
XX
DT      24-APR-2003 (first entry)
XX
DE      Corn ear-derived polynucleotide (cpd) #463.
XX
KW      Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;
KW      structural gene; functional gene; regulatory gene;
KW      corn ear-specific profile; gene transcription; gene expression;
KW      hybrid plant; desirable trait expression; plant breeding program;
KW      inheritance; desired characteristic; growth; development;
KW      disease resistance; environmental adaptability; quality; yield;
KW      multigene trait; plant; gene; ss.
XX
OS      Zea mays.
XX
PN      US6476212-B1.
XX
PD      05-NOV-2002.
XX
PF      14-MAY-1999; 99US-00313294.
XX
PR      26-MAY-1998; 98US-0086722P.
XX
PA      (INCY-) INCYTE GENOMICS INC.

```

XX Lalgudi RV, Ito LY, Sherman BK;  
 XX WPI; 2003-208840/20.  
 XX Novel purified corn-ear derived polynucleotide useful as hybridization  
 PT probe for detecting polynucleotide in sample, and for identifying,  
 PT evaluating, and altering desired characteristics associated with growth,  
 PT development.  
 XX Example; SEQ ID NO 463; 390bp; English.  
 XX The present invention relates to the isolation of corn ear-derived  
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022  
 CC and SATMON023. Some of the cdps uniquely identify structural, functional,  
 CC and regulatory genes of corn ear. The polynucleotide sequences are  
 CC useful for detecting cdps in a sample, for producing a corn ear-specific  
 CC profile of gene transcription, for detecting altered gene expression in  
 CC inbred or hybrid plants, and for screening several molecules for specific  
 CC binding to the polynucleotide. The cdps are useful to identify, isolate,  
 CC or extend identical or related corn-ear nucleic acid sequences from DNA  
 CC libraries, and in nucleic acid amplification or hybridisation techniques  
 CC to follow the expression of desirable traits through plant breeding  
 CC programs. Preferably, the cdps are used to identify, evaluate, alter, or  
 CC follow the inheritance of desired characteristics associated with growth,  
 CC and development, disease resistance, environmental adaptability, quality,  
 CC and yield of corn. The cdps are also useful as molecular markers for  
 CC studying inheritance and multigene traits in a plant breeding program.  
 CC The cdps are useful for producing purified corn-ear polypeptides by  
 CC recombinant techniques. They are also useful in diagnostic assays to  
 CC detect or confirm conditions or diseases associated with abnormal levels  
 CC of cdp expression. ABX81541-ABX89140 represent corn ear-derived  
 CC polynucleotides (cdps) of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/psipdIDEntry.html  
 XX Sequence 274 BP; 83 A; 49 C; 73 G; 67 T; 0 U; 2 Other;  
 SQ  
 Query Match 2.2%; Score 72.2; DB 7; Length 274;  
 Best Local Similarity 60.6%; Pred. No. 2.7e-07;  
 Matches 152; Conservative 0; Mismatches 95; Indels 4; Gaps 2;  
 QY 1777 TATGAGGCTTTAGAGCAGCGCATTCCTATGGT-CCACAGGCGCATCGGAGATGAGT 1835  
 DB 2 TATGAGGCAAGTAMGACGATGCTATGCTGCTCTGGGCGACCGTGNATAGAGCT 61  
 QY 1836 TCGATGTTTGAAGAGCAGTGCATGCTATTGAGGCGCAACGCTCCACCGGAGGT 1895  
 DB 62 GTTATATTTGAAGCTCAGCTGTGGGCTACATGAGAGCTGAACGTGCATMAACACTT 121  
 QY 1896 AGCTGAGATGGGTTATGATGATTAATTCCTGG--GCTCAGAACGCGATATGTTCTGG 1952  
 DB 122 TGTATATCAAGGATACAGACAGAAATTCATGCGACCTACCAAGGTTCATTTGTGCTGG 181  
 QY 1953 AGTGTGTCGCAACGTATGAGCTCTCTGCAACGAAAGCATCTGACATATTCATCA 2012  
 DB 182 TGGGAAAGGCACTATATGTTTCTTAGCAACAAAGAGATATGAGGCAATTAAACA 241  
 QY 2013 ACACCTCTCAAG 2023  
 DB 242 GCATTGCCAAG 252  
 RESULT 5  
 ABL33697 standard; DNA; 6668 BP.  
 XX ABL33697;  
 AC 26-MAR-2002 (first entry)  
 DT Human immune system associated gene SEQ ID NO: 1670.  
 XX

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytosinatic; nocrotropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antineumatic; antiarthritic; antidiabetic; antiporotic;  
 KW antineumatic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
 KW ds.  
 XX Homo sapiens.  
 OS  
 XX WO200200928-A2.  
 PN  
 XX 03-JUN-2002.  
 PD  
 XX 02-JUL-2001; 2001WO-EP007537.  
 PF  
 XX 30-JUN-2000; 2000DE-01032529.  
 PR  
 XX 01-SEP-2000; 2000DE-01043826.  
 PR  
 XX (EPIC-) EPIDENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX WPI; 2002-130909/17.  
 DR  
 PT Nucleic acid comprising fragment of chemically modified gene, useful for  
 PT diagnosis and treatment of diseases associated with abnormal cytosine  
 PT methylation.  
 PT  
 XX Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.  
 PS  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention  
 CC  
 SQ Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 0 U; 1 Other;  
 Query Match 2.1%; Score 68.8; DB 6; Length 6668;  
 Best Local Similarity 45.8%; Pred. No. 7.6e-06;  
 Matches 238; Conservative 0; Mismatches 282; Indels 0; Gaps 0;  
 QY 156 GTTTTGAAGCTTAATAAGCTTCTCTATGCTCTTCTGCTAGTTATTTCTTCCTC 215  
 DB 2742 GTTTTGTGTTTGT 2801  
 QY 216 CGAGTCTGACACACACCTGCTCCGCGGCTTAAGCTTACGTTCCGTCGTT 275  
 DB 2802 TTTTGTGTTTGT 2861  
 QY 276 ACTCTGAAGTTTTCGCTTGAAGCCTCGAGTGCCTGACCGAGCATTTCTGCTCG 335  
 DB 2862 TTTTGTGTTTGT 2921  
 QY 336 ATTCTCTTTTCTGCTGGAAGAAATGCGCCTATGTCGATTTGGAAGTTTGT 395  
 DB 2922 TTTTGTGTTTGT 2981  
 QY 396 GCTATGGGTACTTTTTCCTATATTTTATGTTCTTAGTAAAGATCTGCGTCTA 455  
 DB 2962 TTTTGTGTTTGT 3041  
 QY 456 CTGTTTGTGTTTCTGTTGCTGTTGCTTACCGTTAGTGTGAGATTTGACTG 515  
 DB 3042 TTTTGTGTTTGT 3101  
 QY 516 TGAAATATCCCTTCGTTTGTGTTTGTTCATATAAATGAGATGATACCTTTTGT 575

[illegible]

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
SQ

Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

Query Match 2.1%; Score 68.6; DB 5; Length 556;  
Best Local Similarity 48.3%; Pred. No. 2.8e-06;  
Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
265 CTCGCGTCTTACTCTGTAAGTTTCTGCTTAGAGCCCTCGATGCGCTCACCGCATGCA 324
466 CCCCCCTCCCCCTTTTATATTTTATTTTCTCCCCCCCCCTCCCATTAATA 427
325 TTCTGTGCTGATTTCTCTTTCTCTGCGTGAATAATGCGCTAATGCTCGATTTGCG 384
426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 367
365 AAGTTTCTGCTAGTGGTCTTTTCTTCCCTAATTTTATAGTTCTTAGTAACGATA 444
366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307
445 CCGCGCTTACTCTGTAAGTTTCTGCTTAGAGCCCTCGATGCGCTCACCGCATGCA 504
306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247
505 GATTTGACGTGGAATAATCCTGCTTTTGTGTTGTTTGTTCATATAAATCGATTTGAT 564
246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187
565 CTACCTTTGCTGCTTGAATTTGTTTGTGAGCCATGCGCTTGTGCTGTATAGTACT 624
186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127
625 TCACGTTCAATGTGATTTTGAGATTTT 653
126 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 98
```

## RESULT 8

ABV42105/C

ID ABV42105 standard; cDNA; 556 BP.

XX AC ABV42105;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 42096.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219007P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JB;

XX DR WPI; 2001-662795/76.

XX PR Novel isolated nucleic acid molecule associated with cancerous state of  
XX PT prostate cells and correlating with presence of prostate cancer, useful  
XX PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 8441; 11750bp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX specification or its complement. (I) is useful for: (a) assessing whether  
XX a patient is afflicted with prostate cancer; (b) monitoring the efficacy  
XX progression of prostate cancer in a patient; (c) assessing the efficacy  
XX or a test compound to inhibit prostate cancer in a patient; (d) assessing  
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;  
XX (e) selecting a composition for inhibiting prostate cancer in a patient;  
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
XX determining whether prostate cancer has metastasized in a patient; (h)  
XX assessing the aggressiveness or indolence of prostate cancer in a patient  
XX ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX

SQ Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

Query Match 2.1%; Score 68.6; DB 5; Length 556;  
Best Local Similarity 48.3%; Pred. No. 2.8e-06;  
Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
265 CTCGCGTCTTACTCTGTAAGTTTCTGCTTAGAGCCCTCGATGCGCTCACCGCATGCA 324
486 CCCCCCTCCCCCTTTTATATTTTATTTTCTCCCCCCCCCTCCCATTAATA 427
325 TTCTGTGCTGATTTCTCTTTCTCTGCGTGAATAATGCGCTAATGCTCGATTTGCG 384
426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 367
385 AAGTTTCTGCTAGTGGTCTTTTCTTCCCTAATTTTATAGTTCTTAGTAACGATA 444
366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307
445 CCGCGCTTACTCTGTAAGTTTCTGCTTAGAGCCCTCGATGCGCTCACCGCATGCA 504
306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247
505 GATTTGACGTGGAATAATCCTGCTTTTGTGTTGTTTGTTCATATAAATCGATTTGAT 564
246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187
565 CTACCTTTGCTGCTTGAATTTGTTTGTGAGCCATGCGCTTGTGCTGTATAGTACT 624
186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127
625 TCACGTTCAATGTGATTTTGAGATTTT 653
126 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 98
```

## RESULT 9

ABV43601/C

ID ABV43601 standard; cDNA; 556 BP.

XX AC ABV43601;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 43592.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX 23-AUG-2001.  
PD 20-FEB-2001; 2001WO-US005171.  
XX  
PF 17-FEB-2000; 2000US-0183319P.  
XX 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PI Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
PS Claim 1; Page 8681-8682; 11750PP; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC assessing whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;  
Query Match  
Best Local Similarity 2.1%; Score 68.6; DB 5; Length 556;  
Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;  
QY 265 CTCGCTGTTTACCTGTAAGTTTCTGCTTAGAGCCTCGATCGCTACCGCATGCA 324  
Db 486 CNGCCTCTCCCTTTTAAATATTTTATTTTCTCCCCCTCCCAATAAATA 427  
QY 325 TTCTGCTGATTTCTTTTCTGCTGGAATAATGCTTAATGTTCTCGATTTCG 384  
Db 426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 367  
QY 385 AAGGTTTGTGCTAAGGTTACTTTTCCCTAATTTATTAAGTTCTTAAGCAATA 444  
Db 366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 307  
QY 445 CTGCGCTACTAGTTTGTGTAATTTGTTGCTTCAACCGTTAGTGCCTAATGGA 504  
Db 306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 247  
QY 505 GATTTGACTGTAATAATCTTCGTTTGTGTTTGTGTTTCAATATAATCGATTGAT 564  
Db 246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 187  
QY 565 CTACCTTTGTGCTGATGTTGTTTGTGAGCCATAGCGTTGCTGTTAATACT 624  
Db 186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 127  
QY 625 TCACGTTCACTGATGATTTGAGATTT 653  
Db 126 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 98  
RESULT 10

AA545347  
ID AA545347 standard; DNA; 9539 BP.  
XX  
AC AA545347;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Chemically pretreated complementary DNA associated with cell cycle #26.  
XX  
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;  
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
KW PCR primer.  
XX  
OS Homo sapiens.  
XX  
XX WO200168911-A2.  
XX  
XX 20-SEP-2001.  
XX  
XX 15-MAR-2001; 2001WO-EP002945.  
XX  
XX 15-MAR-2000; 2000DE-01013847.  
XX 06-APR-2000; 2000DE-01019058.  
XX 07-APR-2000; 2000DE-01019173.  
XX 30-JUN-2000; 2000DE-01032529.  
XX 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPig-) EPiGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-602751/68.  
XX  
XX Designing primers and probes for analyzing diseases associated with  
PT cytosine methylation state e.g. arthritis, cancer, aging,  
PT arteriosclerosis comprising fragments of chemically modified genes  
PT associated with cell cycle.  
XX  
PS Claim 1; SEQ ID NO 52; 28pp; English.  
XX  
XX Sequences AA545296-AA545520 represent chemically pretreated genomic DNA  
CC molecules associated with the cell cycle and specific PCR primers of the  
CC invention. The sequences are useful for detecting the methylation state  
CC of all CpG dinucleotides in a sequence and therefore for analyzing  
CC associated diseases. By analysing cytosine methylations in the pretreated  
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
CC of existing diseases or the predisposition to specific diseases can be  
CC ascertained. The parameters may be compared to another set of genetic  
CC and/or epigenetic parameters, the differences serving as basis for  
CC diagnosis and/or prognostic events which are disadvantageous to patients.  
CC The sequences of the invention are useful for the diagnosis and therapy  
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers  
XX  
SQ Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;  
Query Match  
Best Local Similarity 2.0%; Score 66.8; DB 4; Length 9539;  
Matches 212; Conservative 0; Mismatches 242; Indels 0; Gaps 0;  
QY 251 TTTAACTTACGTTCTCCGCTTTTACTCTGTAAGTTTCTGCTAGACCTCCGATCG 310  
Db 202 TTTTCTTTTGTATTTTTCGTTTATTTTGTGCTTTTCTTTTCTTTTCTTTTCT 261  
QY 311 CCTACCGCATGATCTGTGCTGATTTCTCTTTTCTTGTGCTGGAATAATGCCCTAA 370  
Db 262 GTTTTCTTTTCTTTTGTGCTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 321  
QY 371 TGTTCTGATTTGGAAGTTTCTTGTGCTAATGCTTACTTTTCCCTAATATTTAAGTT 430







KM Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;  
 KM trichiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.  
 XX  
 OS unidentified.  
 XX  
 PN WO200181622-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 06-APR-2001; 2001WO-EP003972.  
 XX  
 PR 06-APR-2000; 2000DE-01019058.  
 PR 07-APR-2000; 2000DE-01019173.  
 PR 30-JUN-2000; 2000DE-01032529.  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX  
 PA (EPIC-) EPIDENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K;  
 PI  
 DR WPI; 2002-034446/04.  
 XX  
 PT New nucleic acid derived from genes associated with DNA repair, useful  
 PT for diagnosis, e.g. of ataxia telangiectasia, by determination of  
 PT cytosine methylation.  
 XX  
 PS Claim 1; SEQ ID NO 122; 25pp + Sequence Listing; English.  
 XX  
 CC The invention relates to nucleic acids containing a sequence of at least  
 CC 18 nucleotides of chemically treated DNA of genes associated with DNA  
 CC repair, and their complements. The invention also relates to nucleic  
 CC acids comprising at least 18 base pairs of the chemically pretreated DNA  
 CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,  
 CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, NUDT1, TIG, INPPL1,  
 CC RFC4, DDTL, FANCB, or XRCC8. Nucleic acids of the invention and related  
 CC oligomers, are useful for diagnosis of diseases associated with gene  
 CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,  
 CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,  
 CC immunodeficiency, trichiodystrophy, Fanconi's anaemia, solid tumours  
 CC and cancer, particularly by determining status of cytosine methylation  
 CC and/or by detecting single-nucleotide polymorphisms. Determination of  
 CC individual methylation patterns may allow development of individualised  
 CC therapies. The sequences given in records ABL92192-ABL92335 represent  
 CC chemically pre-treated DNA fragments from genes associated with DNA  
 CC repair, and their complements. Note: The sequence data for this patent is  
 CC not represented in the specification, but is based on sequence  
 CC information supplied by the European Patent Office  
 XX  
 SO Sequence 8079 BP; 1992 A; 212 C; 1962 G; 3913 T; 0 U; 0 Other;  
 Query Match 2.0%; Score 66.4; DB 6; Length 8079;  
 Best Local Similarity 48.2%; Pred. No. 3.2e-05;  
 Matches 187; Conservative 0; Mismatches 201; Indels 0; Gaps 0;  
 QY 333 TCGATTCCTCTTTCTGCGTGGAAAAATGCCCTATGTTTCGATTTGGAAGTTT 392  
 Db 5327 TTGTTGATGTTTTTATATATATAGTGAAGTGCTTATGTTGTTTGTGTTTTT 5386  
 QY 393 TGGCTAAGGCTACTTTTCCCTAATTTTATGTTCTAGGTAACGATACCTGCGTC 452  
 Db 5387 TT 5446  
 QY 453 TTACTGTTTTGTCATTTTGTGCTTACCGCTTAGCGCTGATCGAGATTGA 512  
 Db 5447 TTTTTTTTTTTTGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5506  
 QY 513 CTGGAATAAATCCCTGTTTTGTTTTGTTTCATATAAATGATGATACCTT 572  
 Db 5507 TT 5566  
 QY 573 TGGCTTGTATGTTGTTTTTGAAGCTATGCGTTGCTGTTATTAACCTGACGTC 632  
 Db 5567 TT 5626

QY 633 ATGTGTGATTTTGAGATTTTGATGACTGTGGGTTTTTGTGGCTATAGTTGTA 692  
 Db 5627 TTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAGATGCTGTTTGTGTTATTTAGCTGGA 5686  
 QY 693 AAAATGAGTTCTAGGCGCTGTCATGTT 720  
 Db 5687 GTGTAGTGTATGTTATGTTATGTTATGTT 5714  
 RESULT 13  
 AAX33181/C  
 ID AAX33181 standard; DNA; 6644 BP.  
 XX  
 AC AAX33181;  
 XX  
 DT 25-JUN-1999 (first entry)  
 XX  
 DE Base sequence of the plasmid pRX-ires-bar.  
 XX  
 KM Cowpox virus; bar; viral vector; expression; apoptosis; resistance; crmA;  
 KM bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;  
 KM autoimmune disease; graft rejection reaction; inflammation;  
 KM inflammatory disease; ss.  
 XX  
 OS Synthetic.  
 OS Cowpox virus.  
 XX  
 PN WO9913073-A2.  
 XX  
 PD 18-MAR-1999.  
 XX  
 PR 07-SEP-1998; 98WO-JP004010.  
 XX  
 PR 08-SEP-1997; 97JP-00259235.  
 XX  
 PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.  
 PI Hamada H;  
 PI  
 DR WPI; 1999-243728/20.  
 XX  
 PT New apoptosis-resistant virus-sensitive cell.  
 PT  
 PS Example 1; Page 38-41; 51pp; English.  
 XX  
 CC The present invention describes an apoptosis-resistant virus-sensitive  
 CC cell line into which an apoptosis resistance gene has been introduced.  
 CC The recombinant viruses generated are capable of expressing apoptosis-  
 CC associated genes. These can then be used in a variety of diseases for  
 CC which the induction of apoptosis by gene transfer, or where the  
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
 CC are useful as vectors for gene therapy which can be applied to cancer  
 CC therapy for destroying cancer cells selectively, the treatment of  
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
 CC encountered the problem where if an adenovirus vector capable of  
 CC expressing an apoptosis-associated gene is introduced into animal cells,  
 CC the cells producing the virus will be destroyed because the period of  
 CC time required to induce cell death by apoptosis is shorter than that  
 CC required to replicate and produce the virus, resulting in failure to  
 CC obtain a recombinant virus having the integrated apoptosis-associated  
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
 CC apoptosis resistant gene introduced) is established and overcomes the  
 CC problem. The present sequence represents the base sequence of the plasmid  
 CC pRX-ires-bar, which contains the cowpox virus bar gene, and is used in an  
 CC example from the present invention  
 XX  
 SO Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 U; 0 Other;  
 Query Match 2.0%; Score 66; DB 2; Length 6644;  
 Best Local Similarity 45.4%; Pred. No. 3.6e-05;  
 Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0;





**This Page Blank (uspto)**



```
US-10-311-455-1214
; Sequence 1214, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1214
; LENGTH: 9888
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1214

Query Match
Best Local Similarity 0.7%; Score 22; DB 14; Length 9888;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3142 TAGATTTTGGTTTGGT 3163
DB 8219 TAGATTTTGGTTTGGT 8240

RESULT 3
US-10-027-632-250839
; Sequence 250839, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250839
; LENGTH: 2164
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250839

Query Match
Best Local Similarity 0.6%; Score 21; DB 15; Length 2164;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3154 TTTTGGTAAATTTCATA 3174
DB 842 TTTTGGTAAATTTCATA 862

RESULT 4
US-10-311-455-482
; Sequence 482, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 482
; LENGTH: 6048
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-482

Query Match
Best Local Similarity 0.6%; Score 21; DB 14; Length 6048;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 526 TTGGTTTGGTTTGGTT 546
DB 5395 TTGGTTTGGTTTGGTT 5415

RESULT 5
US-10-311-455-4
; Sequence 4, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 4
; LENGTH: 6072
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-4

Query Match
Best Local Similarity 0.6%; Score 21; DB 14; Length 6072;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```





Best Local Similarity 100.0%; Pred. No. 32,  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3164 AAAATTCATATGAAAGTTA 3183  
|||||

Db 153 AAAATTCATATGAAAGTTA 134  
|||||

## RESULT 10

US-10-074-475-133  
; Sequence 133, Application US/10074475  
; Publication No. US20030092898A1  
; GENERAL INFORMATION:  
; APPLICANT: Saiceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Hu, Ping  
; APPLICANT: Reardon, Hervé  
; APPLICANT: Karra, Kalpana  
; APPLICANT: Caferkey, Robert  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific  
; FILE REFERENCE: DEK-0313  
; CURRENT APPLICATION NUMBER: US/10/074,475  
; PRIOR FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 60/268,292  
; NUMBER OF SEQ ID NOS: 295  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 133  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-074-475-133

Query Match  
Best Local Similarity 100.0%; Score 20; DB 14; Length 1023;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3136 TTTCCTAGTATTTGTTT 3155  
|||||

Db 832 TTTCCTAGTATTTGTTT 851  
|||||

## RESULT 11

US-10-424-599-91177  
; Sequence 91177, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovacic, David K  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO: 91177  
; LENGTH: 1377  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53343C.1  
US-10-424-599-91177

Query Match  
Best Local Similarity 100.0%; Score 20; DB 12; Length 1377;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 CAGTGCATTGTCACGCTTG 1387

Db 359 CAGTGCATTGTCACGCTTG 378  
|||||

## RESULT 12

US-10-282-122A-10949/c  
; Sequence 10949, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haeelbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zykkind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIOR FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See file wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 10949  
; LENGTH: 1461  
; TYPE: DNA  
; ORGANISM: Borrelia burgdorferi  
US-10-282-122A-10949

Query Match  
Best Local Similarity 100.0%; Score 20; DB 12; Length 1461;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2047 ATATATGCTTTTACTTTGT 2066  
|||||

Db 1073 ATATATGCTTTTACTTTGT 1054  
|||||

## RESULT 13

US-10-424-599-91176  
; Sequence 91176, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovacic, David K  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91176
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
US-10-424-599-91176

```

```

Query Match          0.6%; Score 20; DB 12; Length 1589;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1368 CAGTGCATTGTCACGCTTG 1387
          |||||
Db      359 CAGTGCATTGTCACGCTTG 378

```

```

RESULT 14
US-10-108-260A-1756
; Sequence 1756, Application US/10108260A
; Publication No. US2004005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US2004005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1756
; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1756

```

```

Query Match          0.6%; Score 20; DB 15; Length 2418;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3165 AAATTTCATTGGAAGTTAG 3184
          |||||
Db      1988 AAATTTCATTGGAAGTTAG 2007

```

```

RESULT 15
US-09-947-953-1/C
; Sequence 1, Application US/09947953
; Patent No. US2002015101A1
; GENERAL INFORMATION:
; APPLICANT: DONAHUE, J. KEVIN
; APPLICANT: MARBAN, EDUARDO
; TITLE OF INVENTION: CARDIAC ARRHYTHMIA TREATMENT METHODS
; FILE REFERENCE: 71699/56415
; CURRENT APPLICATION NUMBER: US/09/947,953
; CURRENT FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,311
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/295,889
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2985
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-947-953-1

```

```

Query Match          0.6%; Score 20; DB 9; Length 2985;

```

```

Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      71 CGAAAAAGGAAAAAAGG 90
          |||||
Db      25 CGAAAAAGGAAAAAAGG 6

```

```

Search completed: March 30, 2004, 23:20:26
Job time : 2047.85 secs

```

**This Page Blank (uspto)**



TYPE: DNA  
ORGANISM: Escherichia coli  
US-09-736-734-1

Query Match 0.6%; Score 20; DB 4; Length 654;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3164 AAAATTCATATGAAGTTA 3183  
|||||  
Db 153 AAAATTCATATGAAGTTA 134

RESULT 3  
US-09-615-192A-109/c  
Sequence 109, Application US/09615192A  
Patent No. 6410718  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003c4U  
CURRENT FILING DATE: 2000-07-12  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1996-09-11  
PRIOR APPLICATION NUMBER: US 09/169,789  
PRIOR FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 405  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 109  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-615-192A-109

Query Match 0.6%; Score 19; DB 4; Length 481;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAGGTTTGGCAGG 787  
|||||  
Db 470 TGGTTCAGGTTTGGCAGG 452

RESULT 4  
US-09-169-789-109/c  
Sequence 109, Application US/09169789  
Patent No. 6653528  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003c2  
CURRENT FILING DATE: 1998-10-09  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1996-09-11  
NUMBER OF SEQ ID NOS: 185  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 109  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-169-789-109

Query Match 0.6%; Score 19; DB 4; Length 481;

Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 769 TGGTTCAGGTTTGGCAGG 787  
|||||  
Db 470 TGGTTCAGGTTTGGCAGG 452

RESULT 5  
US-08-975-316-27/c  
Sequence 27, Application US/08975316  
Patent No. 5952486  
GENERAL INFORMATION:  
APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka  
TITLE OF INVENTION: MATERIALS AND METHODS FOR  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,316  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/713,000  
FILING DATE: September 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: SLEATH, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000/1003C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 495 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-975-316-27

Query Match 0.6%; Score 19; DB 2; Length 495;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAGGTTTGGCAGG 787  
|||||  
Db 484 TGGTTCAGGTTTGGCAGG 466

RESULT 6  
US-09-615-192A-27/c  
Sequence 27, Application US/09615192A  
Patent No. 6410718  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003c4U  
CURRENT APPLICATION NUMBER: US/09/615,192A

CURRENT FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: US 08/713,000  
PRIOR FILING DATE: 1996-09-11  
PRIOR APPLICATION NUMBER: US 09/169,789  
PRIOR FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 405  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 27  
LENGTH: 495  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-615-192A-27

Query Match 0.6%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAAGTTTGGCAGG 787  
|||||  
DB 484 TGGTTCAAGTTTGGCAGG 466

RESULT 7  
US-09-169-789-27/c  
Sequence 27, Application US/09169789  
Patent No. 6653528  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003C2  
CURRENT APPLICATION NUMBER: US/09/169,789  
CURRENT FILING DATE: 1998-10-09  
EARLIER APPLICATION NUMBER: US 08/975,316  
EARLIER FILING DATE: 1997-11-21  
EARLIER APPLICATION NUMBER: US 08/713,000  
EARLIER FILING DATE: 1996-09-11  
NUMBER OF SEQ ID NOS: 185  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 27  
LENGTH: 495  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-169-789-27

Query Match 0.6%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAAGTTTGGCAGG 787  
|||||  
DB 484 TGGTTCAAGTTTGGCAGG 466

RESULT 8  
US-08-484-105-19  
Sequence 19, Application US/08484105  
Patent No. 5589341  
GENERAL INFORMATION:  
APPLICANT: STILLMAN, Bruce  
APPLICANT: BELL, Stephen P  
APPLICANT: KOBAYASHI, Ryuji  
APPLICANT: RINE, Jasper  
APPLICANT: FOSS, Margit  
APPLICANT: MCNALLY, Francis J  
APPLICANT: LAURENSEN, Patricia  
APPLICANT: HERSKOWITZ, Ira  
APPLICANT: LI, Joachim J  
APPLICANT: GAVIN, Kimberly  
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,105  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1480 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 277..1365  
US-08-484-105-19

Query Match 0.6%; Score 19; DB 1; Length 1480;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1655 TAAAGTGAATTCCTCTGT 1673  
|||||  
DB 238 TAAAGTGAATTCCTCTGT 256

RESULT 9  
US-08-484-106-19  
Sequence 19, Application US/08484106  
Patent No. 5614618  
GENERAL INFORMATION:  
APPLICANT: STILLMAN, Bruce  
APPLICANT: BELL, Stephen P  
APPLICANT: KOBAYASHI, Ryuji  
APPLICANT: RINE, Jasper  
APPLICANT: FOSS, Margit  
APPLICANT: MCNALLY, Francis J  
APPLICANT: LAURENSEN, Patricia  
APPLICANT: HERSKOWITZ, Ira  
APPLICANT: LI, Joachim J  
APPLICANT: GAVIN, Kimberly  
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

RESULT 13  
 US-09-595-684B-30  
 Sequence 30, Application US/09595684B  
 Patent No. 6544766  
 GENERAL INFORMATION:  
 APPLICANT: Beraud, Christophe  
 APPLICANT: Ohashi, Cara  
 APPLICANT: Sakowicz, Roman  
 APPLICANT: Vaisberg, Eugene  
 APPLICANT: Wood, Kenneth  
 APPLICANT: Yu, Ming  
 TITLE OF INVENTION: Human kinesins and methods of producing  
 TITLE OF INVENTION: and purifying human kinesins  
 FILE REFERENCE: cytopo036  
 CURRENT APPLICATION NUMBER: US/09/595,684B



```

; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 8257
; TYPE: DNA
; ORGANISM: Human
US-09-595-684B-30

```

```

Query Match 0.6%; Score 19; DB 4; Length 8257;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2873 TTGATCTGAGAGAAAGATT 2891
DB 2141 TTGATCTGAGAGAAAGATT 2159

```

#### RESULT 14

```

US-09-620-312D-130
; Sequence 130, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE REFERENCE: 784C1P2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_fl_genes Version 1.0
; SEQ ID NO 130
; LENGTH: 8503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (91)..(8082)
US-09-620-312D-130

```

```

Query Match 0.6%; Score 19; DB 4; Length 8503;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2873 TTGATCTGAGAGAAAGATT 2891
DB 2141 TTGATCTGAGAGAAAGATT 2159

```

RESULT 15  
US-09-163-748C-3/c

```

; Sequence 3, Application US/09163748C
; Patent No. 6509172
; GENERAL INFORMATION:
; APPLICANT: Debacker, Oliver
; APPLICANT: Van den Eynde, Benoit
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode Member Of The Gage
; FILE REFERENCE: IUD 5558
; CURRENT APPLICATION NUMBER: US/09/163,748C
; CURRENT FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 27
; SEQ ID NO 3
; LENGTH: 9531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 92,232,1041,7412,9038-9040
; OTHER INFORMATION: Identity of several nucleotides not known
US-09-163-748C-3

```

```

Query Match 0.6%; Score 19; DB 4; Length 9531;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 91 TACAAAATGAAAACAAA 109
DB 7045 TACAAAATGAAAACAAA 7027

```

Search completed: March 30, 2004, 18:11:43  
Job time : 250.688 secs

**This Page Blank (uspto)**



XX Beclin C, Elmayan T, Vaucheret H;  
 XX  
 DR WPI; 2001-159529/16.  
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 PT resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 XX  
 PS Claim 1; Page 31-32; 36pp; French.  
 XX  
 CC The present sequence represents the genomic sequence of the Arabidopsis  
 CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional  
 CC inactivation (degradation of RNA) and for resistance to viruses.  
 CC Overexpression of SGS3 results in plants with increased resistance to  
 CC viruses; while inactivation of SGS3 in transgenic plants (e.g. by  
 CC expressing antisense RNA, by mutation or by homologous recombination)  
 CC increases the level of the transgene product. This product may e.g.  
 CC impart resistance (to herbicide, insects or pathogens), alter contents of  
 CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an  
 CC immunoglobulin or interferon  
 CC  
 SQ Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 3275; DB 4; Length 3275;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAACAAACAAATTTAGCAAGTCATGTCGTAGCAATTAATTAATAGTGGACAA 60  
 DB 1 GACAAACAAACAAATTTAGCAAGTCATGTCGTAGCAATTAATTAATAGTGGACAA 60  
 QY 61 TTAAGTTAAGCGAAAAAGAAAAAAGGTACAAAAATGAAAAAATAAACTGAT 120  
 DB 61 TTAAGTTAAGCGAAAAAGAAAAAAGGTACAAAAATGAAAAAATAAACTGAT 120  
 QY 121 GAAAAATTGAGTCGAAATCGAAAAAGAGCGGTTTAAAGCTTAATAGCTTCTC 180  
 DB 121 GAAAAATTGAGTCGAAATCGAAAAAGAGCGGTTTAAAGCTTAATAGCTTCTC 180  
 QY 181 ATTGTCCTCTGTCGAGTTTATTTCTTCTCCGAGTCGACCTGACCTACCTCAG 240  
 DB 181 ATTGTCCTCTGTCGAGTTTATTTCTTCTCCGAGTCGACCTGACCTACCTCAG 240  
 QY 241 TCTCCGCGCTTTAAACTTACGTTCTCCGTCGTTACTCTGTAAGTTTCTGCTTGA 300  
 DB 241 TCTCCGCGCTTTAAACTTACGTTCTCCGTCGTTACTCTGTAAGTTTCTGCTTGA 300  
 QY 301 CCTCGATCGCTCACCAGATGCAATTCGTGCTGCAATTCCTTTTCTGCTGAAAA 360  
 DB 301 CCTCGATCGCTCACCAGATGCAATTCGTGCTGCAATTCCTTTTCTGCTGAAAA 360  
 QY 361 ATGCGCTAATGTTCTGATTTGCAAGGTTTGTGCAATGGGTAATTTTCCCTATA 420  
 DB 361 ATGCGCTAATGTTCTGATTTGCAAGGTTTGTGCAATGGGTAATTTTCCCTATA 420  
 QY 421 TTTTAATGTTCTTAAAGTACATACCTGCTTACTGTTTGTGCAATTTGTTGCT 480  
 DB 421 TTTTAATGTTCTTAAAGTACATACCTGCTTACTGTTTGTGCAATTTGTTGCT 480  
 QY 481 TTTACCGGTTTGTGCTGATGAGATTTTGAATGCTGAAAAATCTCTGTTTGTG 540  
 DB 481 TTTACCGGTTTGTGCTGATGAGATTTTGAATGCTGAAAAATCTCTGTTTGTG 540  
 QY 541 TTGTTTCATATAATACGATGATCTTGTGCTGTTGATGTTGTTTGTGACCT 600  
 DB 541 TTGTTTCATATAATACGATGATCTTGTGCTGTTGATGTTGTTTGTGACCT 600  
 QY 601 ATGCGTTGTTGGCTTTGTAATCACTTCAGTTCAATGTTGATTTTGGTGGT 660  
 DB 601 ATGCGTTGTTGGCTTTGTAATCACTTCAGTTCAATGTTGATTTTGGTGGT 660  
 QY 661 ACTGTGGGTTCTTTGTGCTATAGGTTGTAATAATGATTTCTAGGGCTGCTCAATGT 720

DB 661 ACTGTGGGTTCTTTGTGCTATAGGTTGTAATAATGATTTCTAGGGCTGCTCAATGT 720  
 QY 721 CTAAGAAAAAGACGTTCAAGGTGTAATAGGCTGAGTTGAACGTTGTTCAAGTT 780  
 DB 721 CTAAGAAAAAGACGTTCAAGGTGTAATAGGCTGAGTTGAACGTTGTTCAAGTT 780  
 QY 781 TGGCAAGGACGAGACTGCTTCTTCAAGATGATGAGAGAGTGGAGTCAATTTCCA 840  
 DB 781 TGGCAAGGACGAGACTGCTTCTTCAAGATGATGAGAGAGTGGAGTCAATTTCCA 840  
 QY 841 AGAAGAACAAAGAAACCAAGAAACCTTCTGAAAAAATTTGGTTTCTCAAGATTTCCA 900  
 DB 841 AGAAGAACAAAGAAACCAAGAAACCTTCTGAAAAAATTTGGTTTCTCAAGATTTCCA 900  
 QY 901 ATCTCTTAAGAGCTTGGGCTGTGTCAGCAAGAGGAGGTAGCAAGTATCTGGAGAG 960  
 DB 901 ATCTCTTAAGAGCTTGGGCTGTGTCAGCAAGAGGAGGTAGCAAGTATCTGGAGAG 960  
 QY 961 GAAACAAATGATTCGAGAGGATACGCGCAATGTCGAGGCAATTCAGCTAACATATCTG 1020  
 DB 961 GAAACAAATGATTCGAGAGGATACGCGCAATGTCGAGGCAATTCAGCTAACATATCTG 1020  
 QY 1021 GTGGGGGAGAGGCTTGAAGAGAAATGATTAACAATTTGTGCAACCCCACTGTAT 1080  
 DB 1021 GTGGGGGAGAGGCTTGAAGAGAAATGATTAACAATTTGTGCAACCCCACTGTAT 1080  
 QY 1081 CTGCGCTCTTGTGAAGAGATGGAATTTGCAAGCAAGAGAGGTTCTGCTCAGCACA 1140  
 DB 1081 CTGCGCTCTTGTGAAGAGATGGAATTTGCAAGCAAGAGAGGTTCTGCTCAGCACA 1140  
 QY 1141 CAGCTGTGCAAGGATTTCTGAGCTGAGATGATGATTAATGCTTCTGAGAAAGAGA 1200  
 DB 1141 CAGCTGTGCAAGGATTTCTGAGCTGAGATGATGATTAATGCTTCTGAGAAAGAGA 1200  
 QY 1201 ATGATTCGATGCTTTGATGATTTGATGATGATGATGATGATGATGATGATGATGAT 1260  
 DB 1201 ATGATTCGATGCTTTGATGATTTGATGATGATGATGATGATGATGATGATGATGAT 1260  
 QY 1261 ATGTGATCAAAAGAGCCATGATCAAGAACAGAAATAGTGTCAAAAAGTTCTTTG 1320  
 DB 1261 ATGTGATCAAAAGAGCCATGATCAAGAACAGAAATAGTGTCAAAAAGTTCTTTG 1320  
 QY 1321 GCACTTGAATAGCTTGTGATCGAGCAATTAATGAACCAAGAGCGATGTC 1380  
 DB 1321 GCACTTGAATAGCTTGTGATCGAGCAATTAATGAACCAAGAGCGATGTC 1380  
 QY 1381 CAGCTTGTCAAGACGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 DB 1381 CAGCTTGTCAAGACGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 QY 1441 ATGCGAGCAAAAGAGCTTAGCGAGTTAAGCTTCATAGAAATTTGGCTGAAGTTTAG 1500  
 DB 1441 ATGCGAGCAAAAGAGCTTAGCGAGTTAAGCTTCATAGAAATTTGGCTGAAGTTTAG 1500  
 QY 1501 AAAAGGATCTACAGATGAGAGGCAATGCTCTTGTGCTGATGATTTATGGGAGT 1560  
 DB 1501 AAAAGGATCTACAGATGAGAGGCAATGCTCTTGTGCTGATGATTTATGGGAGT 1560  
 QY 1561 GGAAGGTTTGGGTGAGATGAAAAATTAATGAATTTGCTGCTCAATGTCATCA 1620  
 DB 1561 GGAAGGTTTGGGTGAGATGAAAAATTAATGAATTTGCTGCTCAATGTCATCA 1620  
 QY 1621 TCATGAATCTGATCTGATTAAGAGCAATTAAGATGAGTGAATTTCTGCTTTTAC 1680  
 DB 1621 TCATGAATCTGATCTGATTAAGAGCAATTAAGATGAGTGAATTTCTGCTTTTAC 1680  
 QY 1681 TCTTTAATTTTCTTCTGCAATTTCTGATGATGATGATGATGATGATGATGATGATGAT 1740  
 DB 1681 TCTTTAATTTTCTTCTGCAATTTCTGATGATGATGATGATGATGATGATGATGATGAT 1740  
 QY 1741 ATGCGCAACAAAGAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1800

Db 1741 ATGGCAACCAAGAGCTGCTGGAATACCTTCACAAGTATGAGGCTTTAGACACGCCAT 1800  
 Qy 1801 TCCTATGCTCCACAGGGCATGCGGGATGAGTGTTCGATGTTTGAGAGAGGCCACT 1860  
 Db 1801 TCCTATGCTCCACAGGGCATGCGGGATGAGTGTTCGATGTTTGAGAGAGGCCACT 1860  
 Qy 1861 GGCATATTTGAGAGCCGCAACCGCTCCACCGGAGTTAGCTGAGATGGGTTAGATGAATT 1920  
 Db 1861 GGCATATTTGAGAGCCGCAACCGCTCCACCGGAGTTAGCTGAGATGGGTTAGATGAATT 1920  
 Qy 1921 GCTTGGGCTCAGAACCGCATGATGTTTCTGAGAGTGTTCGCCAATCTGATAGCTTCTT 1980  
 Db 1921 GCTTGGGCTCAGAACCGCATGATGTTTCTGAGAGTGTTCGCCAATCTGATAGCTTCTT 1980  
 Qy 1981 GCAACGAAGCAAGATCTGACATATTCATCAACACTCTCAAGGTTCTCTCCCAAGA 2040  
 Db 1981 GCAACGAAGCAAGATCTGACATATTCATCAACACTCTCAAGGTTCTCTCCCAAGA 2040  
 Qy 2041 AATTGATATATGCTTTAGTTTGTGATTGGAATTTAAAGTTTGTGTCCTGTTAA 2100  
 Db 2041 AATTGATATATGCTTTAGTTTGTGATTGGAATTTAAAGTTTGTGTCCTGTTAA 2100  
 Qy 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAACAGGCTGAAATTCAGTTG 2160  
 Db 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAACAGGCTGAAATTCAGTTG 2160  
 Qy 2161 AAATCATACCAAGAGATGTTGTAAGAGCTGAGGAGATCTCTGAGCAATCAGCAG 2220  
 Db 2161 AAATCATACCAAGAGATGTTGTAAGAGCTGAGGAGATCTCTGAGCAATCAGCAG 2220  
 Qy 2221 CTGAACACTTTTAAAGAACAGCTCTCAAAAACAGAACAGCAGCAAGCTTGAAGAA 2280  
 Db 2221 CTGAACACTTTTAAAGAACAGCTCTCAAAAACAGAACAGCAGCAAGCTTGAAGAA 2280  
 Qy 2281 TCTCTGGAATTTATGACGAGAGAGCTGCGTAAGAACTGCAAGATTAATCGATGAGAA 2340  
 Db 2281 TCTCTGGAATTTATGACGAGAGAGCTGCGTAAGAACTGCAAGATTAATCGATGAGAA 2340  
 Qy 2341 CAGAGAACTAAGATGACGAGATGAAACAGAGGAAAGATGATATTTTCTTGAATAAT 2400  
 Db 2341 CAGAGAACTAAGATGACGAGATGAAACAGAGGAAAGATGATATTTTCTTGAATAAT 2400  
 Qy 2401 CACAAACTTGACATTTTGTATTAACCTGATTCATGATTTTGAATTTATGTCACAA 2460  
 Db 2401 CACAAACTTGACATTTTGTATTAACCTGATTCATGATTTTGAATTTATGTCACAA 2460  
 Qy 2461 AAAAAGCTGCTGCTGTTGAAGATGATGCAACAGAGGTTTTCATGATTCATCAAA 2520  
 Db 2461 AAAAAGCTGCTGCTGTTGAAGATGATGCAACAGAGGTTTTCATGATTCATCAAA 2520  
 Qy 2521 CAGATCCATGAAGAAGAGAGCGAAAGAGAGAAATTCAGATGTTGCGACAGAGAA 2580  
 Db 2521 CAGATCCATGAAGAAGAGAGCGAAAGAGAGAAATTCAGATGTTGCGACAGAGAA 2580  
 Qy 2581 CGTGCCAAGTTGTCGCGACAGACAGCAACATTAATCCCTCTACCAATGACGATTC 2640  
 Db 2581 CGTGCCAAGTTGTCGCGACAGACAGCAACATTAATCCCTCTACCAATGACGATTC 2640  
 Qy 2641 CGAAAGAGTATATGTAATCTAATCAATTAATCCCTCTGCGGTTTGTTCAAACCTA 2700  
 Db 2641 CGAAAGAGTATATGTAATCTAATCAATTAATCCCTCTGCGGTTTGTTCAAACCTA 2700  
 Qy 2701 AGAGTAACGTAATTAATCCGCTTTGATTTCTTCGAGACCTGAGAGAGTCAAGTTC 2760  
 Db 2701 AGAGTAACGTAATTAATCCGCTTTGATTTCTTCGAGACCTGAGAGAGTCAAGTTC 2760  
 Qy 2761 ATGAGTTTCAAGAGAGAGAGATGAGAGATTTTGTGAAAGAGAGAGATGCTGATAAA 2820  
 Db 2761 ATGAGTTTCAAGAGAGAGAGATGAGAGATTTTGTGAAAGAGAGAGATGCTGATAAA 2820  
 Qy 2821 GATCAAGAGAGAGAGATGAGAGATGAGAGAGATGAGAGAGATGATGATCTG 2880  
 Db 2821 GATCAAGAGAGAGAGATGAGAGATGAGAGAGATGAGAGAGATGATGATCTG 2880

Qy 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTAACAGCATGCGCTTCAATGAA 2940  
 Db 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTAACAGCATGCGCTTCAATGAA 2940  
 Qy 2941 GATGATGAGCAAAAGCTGTGTAACAAGACAGACATTAAGTTCTTTGTTTCTTTTG 3000  
 Db 2941 GATGATGAGCAAAAGCTGTGTAACAAGACAGACATTAAGTTCTTTGTTTCTTTTG 3000  
 Qy 3001 GTATGTCGAAAGTAGAGATCTGAGAGCTCCATTTAAATCTAGAGCAAAATCTAAGGA 3060  
 Db 3001 GTATGTCGAAAGTAGAGATCTGAGAGCTCCATTTAAATCTAGAGCAAAATCTAAGGA 3060  
 Qy 3061 GATTATGATTTATATCTCCATTAATTTTATGAGCGGATCTAAGAGACATTAAGTCTT 3120  
 Db 3061 GATTATGATTTATATCTCCATTAATTTTATGAGCGGATCTAAGAGACATTAAGTCTT 3120  
 Qy 3121 GTGACTAAAACCAAGTTCTCTAGTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 3180  
 Db 3121 GTGACTAAAACCAAGTTCTCTAGTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 3180  
 Qy 3181 TTGACATATTTACCAACGTCAGAGTGAATCAAGATGCGCAATCAATCATGTTT 3240  
 Db 3181 TTGACATATTTACCAACGTCAGAGTGAATCAAGATGCGCAATCAATCATGTTT 3240  
 Qy 3241 AGAATTTATATCTCAAAATATATGCGTCAAAAT 3275  
 Db 3241 AGAATTTATATCTCAAAATATATGCGTCAAAAT 3275  
 RESULT 2  
 AAF25374  
 ID AAF25374 standard; cDNA; 1878 BP.  
 XX AC AAF25374;  
 XX DT 15-MAY-2001 (first entry)  
 XX DE Nucleotide sequence of the Arabidopsis SGS3 polypeptide.  
 XX KW SGS3 gene; post-transcriptional inactivation; RNA degradation;  
 XX KM viral resistance; resistance; fatty acid content; protein content; ss.  
 XX OS Arabidopsis thaliana.  
 XX FH Key  
 XX FT CDS  
 XX FT 1.1878  
 XX FT /\*tag= a  
 XX FT /product= "SGS3"  
 XX PN W0200105951-A2.  
 XX PD 25-JAN-2001.  
 XX PF 13-JUL-2000; 2000MO-FR002052.  
 XX PR 16-JUL-1999; 99PR-00009417.  
 XX PR 26-JAN-2000; 2000FR-00001006.  
 XX XX (AVET ) AVENTIS CROPS SCIENCE SA.  
 XX PA (INRG ) INST NAT RECH AGRONOMIQUE.  
 XX PI Beclin C, Elmayan T, Vaucheret H;  
 XX DR WPI; 2001-159529/16.  
 XX DR P-PSDB; AAB31798.  
 XX XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 XX PT resistance in plants and, when inhibited, for increasing transgene  
 XX PT expression.  
 XX PS Claim 1; Page 32-35; 36pp; French.  
 XX XX

CC The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.  
 CC The SGS3 gene is essential for post-transcriptional inactivation  
 CC (degradation of RNA) and for resistance to viruses. Overexpression of  
 CC SGS3 results in plants with increased resistance to viruses, while  
 CC inactivation of SGS3 in transgenic plants (e.g. by expressing antisense  
 CC RNA, by mutation or by homologous recombination) increases the level of  
 CC the transgene product. This product may e.g. impart resistance (to  
 CC herbicide, insects or pathogens) alter contents of essential fatty acids  
 CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or  
 CC interferon  
 CC

Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 29.4%; Score 963; DB 4; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 696 ATGAGTTCTAGGGCTGTCTCAATGCTTAAGAAAAAAGCCTTCAAGGGTGTATAGGCT 755  
 DB 1 ATGAGTTCTAGGGCTGTCTCAATGCTTAAGAAAAAAGCCTTCAAGGGTGTATAGGCT 60  
 QY 756 GAGGTTGAACGTTGGTTCAGAGTTGGCAGGAGAGACTGGCTTCTTCAAGATGAT 815  
 DB 61 GAGGTTGAACGTTGGTTCAGAGTTGGCAGGAGAGACTGGCTTCTTCAAGATGAT 120  
 QY 816 GAGGAGAGTGGAGGTCATTCCAGAAAGAAACAAACAGAGAAACATCTTGA 875  
 DB 121 GAGGAGAGTGGAGGTCATTCCAGAAAGAAACAAACAGAGAAACATCTTGA 180  
 QY 876 AAAAATTGGTTCTCAGAAATTCGATCTCTAGAGCTTGGGGTGTCTACAGCAAGG 935  
 DB 181 AAAAATTGGTTCTCAGAAATTCGATCTCTAGAGCTTGGGGTGTCTACAGCAAGG 240  
 QY 936 AGAGTAGCAACGTTCTGGAGAGAGAAACATGTTATCCGGAAGGTAACGCAATGCT 995  
 DB 241 AGAGTAGCAACGTTCTGGAGAGAGAAACATGTTATCCGGAAGGTAACGCAATGCT 300  
 QY 996 CGGGGCAATTCAGCTAACATATCTGCTGGGGAGAGAGCTGGAGAGAAATGATATAC 1055  
 DB 301 CGGGGCAATTCAGCTAACATATCTGCTGGGGAGAGAGCTGGAGAGAAATGATATAC 360  
 QY 1056 AACTTTGTGGACCCCACTGTATCTGCGCTCTCTTGAAGAGAGATGGAATGGCAG 1115  
 DB 361 AACTTTGTGGACCCCACTGTATCTGCGCTCTCTTGAAGAGAGATGGAATGGCAG 420  
 QY 1116 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTCTCTGAGAGTGAAGATGAT 1175  
 DB 421 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTCTCTGAGAGTGAAGATGAT 480  
 QY 1176 GTGATATATGCTTGAAGAGAGATGATCCGATGCTTGAATGATTCGATGACGAC 1235  
 DB 481 GTGATATATGCTTGAAGAGAGATGATCCGATGCTTGAATGATTCGATGACGAC 540  
 QY 1236 CTTCAGATGATGATTTATGACTCGAGTGTAGTCAAAAAGCCATGATCAGAAACGAG 1295  
 DB 541 CTTCAGATGATGATTTATGACTCGAGTGTAGTCAAAAAGCCATGATCAGAAACGAG 600  
 QY 1296 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGGATAGCTTGTGATGACACATATAAT 1355  
 DB 601 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGGATAGCTTGTGATGACACATATAAT 660  
 QY 1356 GAACACAGAGGAGAGTGTCTCAGCTTGTGAGAACGAGACCTGGTGCCATGATGAG 1415  
 DB 661 GAACACAGAGGAGAGTGTCTCAGCTTGTGAGAACGAGACCTGGTGCCATGATGAG 720  
 QY 1416 TATAAAGTGAACCTCTTACTAGCTCAGAGAGCAAAAAGAGGTAGGCGAGTTAAGCTC 1475  
 DB 721 TATAAAGTGAACCTCTTACTAGCTCAGAGAGCAAAAAGAGGTAGGCGAGTTAAGCTC 780  
 QY 1476 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGAGTGAAGGCGCATCTGTATT 1535  
 DB 781 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGAGTGAAGGCGCATCTGTATT 840

QY 1536 CTTGTGTGATGATTTATGAGGAGTGAAGGGTTTGGGAGAGATGAATTAATGA 1595  
 DB 841 CTTGTGTGATGATTTATGAGGAGTGAAGGGTTTGGGAGAGATGAATTAATGA 900  
 QY 1596 ATTGTCTGAGCTCCAAATGATCATCATCATGATATAGCTGATTAAGACGATTAAGAT 1655  
 DB 901 ATTGTCTGAGCTCCAAATGATCATCATCATGATATAGCTGATTAAGACGATTAAGAT 960  
 QY 1656 AAG 1658  
 DB 961 AAG 963

RESULT 3

AAFP25372/C

ID AAFP25372 standard; DNA; 27 BP.

XX AAFP25372;

XX 15-MAY-2001 (first entry)

XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.

XX SGS3 gene; post-transcriptional inactivation; RNA degradation;

KW viral resistance; resistance; fatty acid content; protein content;

KW PCR primer; ss.

OS Arabidopsis thaliana.

XX WO200105951-A2.

XX 25-JAN-2001.

XX 13-JUL-2000; 2000WO-FR002052.

XX 16-JUL-1999; 99FR-00009417.

XX 26-JAN-2000; 2000FR-00001006.

XX (AVET ) AVENTIS CROPS SCIENCE SA.

XX (INRG ) INST NAT RECH AGRONOMIQUE.

XX Beclin C, Elmayan T, Vaucheret H;

XX WPI; 2001-159529/16.

XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus

PT resistance in plants and, when inhibited, for increasing transgene

PT expression.

XX Example 1; Page 22; 36pp; French.

XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis

CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-

CC transcripional inactivation (degradation of RNA) and for resistance to

CC viruses. Overexpression of SGS3 results in plants with increased

CC resistance to viruses, while inactivation of SGS3 in transgenic plants

CC (e.g. by expressing antisense RNA, by mutation or by homologous

CC recombination) increases the level of the transgene product. This product

CC may e.g. impart resistance (to herbicide, insects or pathogens), alter

CC contents of essential fatty acids or proteins, or is pharmaceutically

CC active, e.g. an immunoglobulin or interferon

XX

Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;

Query Match 0.8%; Score 27; DB 4; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.055;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2926 GGCCTTCAATGATGATGATGAGAC 2952

DB 27 GGCCTTCAATGATGATGATGAGAC 1

```

RESULT 4
AAF25371
ID AAF25371 standard; DNA; 23 BP.
XX
XX AAF25371;
AC
XX
XX 15-MAY-2001 (first entry)
DE PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
XX SGS3 gene; post-transcriptional inactivation; RNA degradation;
KM viral resistance; resistance; fatty acid content; protein content;
KM PCR primer; ss.
XX
XX Arabidopsis thaliana.
OS
XX WO200105951-A2.
XX
XX 25-JAN-2001.
XX
XX 13-JUL-2000; 2000WO-FR002052.
XX
XX 16-JUL-1999; 99FR-00009417.
XX 26-JAN-2000; 2000FR-00001006.
XX
XX (AVET ) AVENTIS CROPS SCIENCE SA.
PA (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX Beclin C, Elmayan T, Vaucheret H;
XX
XX WPI; 2001-159529/16.
XX
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
PT resistance in plants and, when inhibited, for increasing transgene
PT expression.
XX
XX Example 1; Page 22; 36pp; French.
XX
XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
CC transcriptional inactivation (degradation of RNA) and for resistance to
CC viruses. Overexpression of SGS3 results in plants with increased
CC resistance to viruses, while inactivation of SGS3 in transgenic plants
CC (e.g. by expressing antisense RNA, by mutation or by homologous
CC recombination) increases the level of the transgene product. This product
CC may e.g. impart resistance (to herbicide, insects or pathogens), alter
CC contents of essential fatty acids or proteins, or is pharmaceutically
CC active, e.g. an immunoglobulin or interferon
XX
XX
SQ Sequence 23 BP; 6 A; 4 C; 7 G; 6 T; 0 U; 0 Other;
Query Match 0.7%; Score 23; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 693 AAAATGAGTCTAGGGCTGCTCC 715
DB 1 AAAATGAGTCTAGGGCTGCTCC 23

```

```

XX
XX Homo sapiens.
OS
XX
XX WO200202806-A2.
XX
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-EP007470.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIDEMIOLOGY AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-154757/20.
XX
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligonucleotides,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
XX Claim 1; SEQ ID NO 123; 24pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003600), MRP (NM_004996),
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899 and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 6136 BP; 1661 A; 69 C; 1415 G; 2991 T; 0 U; 0 Other;
Query Match 0.7%; Score 22; DB 6; Length 6136;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3145 TATTGTTTCTTTTGTAA 3166
DB 2122 TATTGTTTCTTTTGTAA 2143

```

```

RESULT 5
ABK40041
ID ABK40041 standard; DNA; 6136 BP.
XX
XX AC ABK40041;
XX
XX 21-MAY-2002 (first entry)
XX
XX Human chemically pretreated gene sequence #62 strand 1.
DE
XX
XX Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KM cytochrome; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KM UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

```

```

RESULT 6
ABN80100
ID ABN80100 standard; DNA; 6301 BP.
XX
XX AC ABN80100;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human chemically modified disease associated gene SEQ ID NO 117.
DE
XX
XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KM heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KM dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;

```

KM antidiabetic; cytostatic; anticonvulsant; ds.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200200927-A2.  
 XX 03-JAN-2002.  
 XX 02-JUL-2001; 2001WO-EP007536.  
 XX 30-JUN-2000; 2000DE-01032529.  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX (EPIC-) EPIDENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K;  
 PI WPI; 2002-130908/17.  
 DR Novel nucleic acid useful for diagnosis and therapy of diseases  
 XX associated with development genes such as diabetes; comprises a sequence  
 PT of a segment of chemically pretreated DNA of genes associated with  
 PT development.  
 XX Claim 1; SEQ ID NO 117; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 16 bases in length of a segment of chemically pretreated DNA (II)  
 CC of genes associated with development selected from 87 genes listed in the  
 CC specification such as ACCPN, ADPN, or APD1 and comprising one of 350  
 CC sequences (ABN79984-ABN80333) or their complements. The invention is  
 CC useful for the diagnosis or therapy of diseases associated with  
 CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Currairio syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all CpG dinucleotides within the 350 sequences or  
 CC (II) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (II) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this  
 CC patent did not form part of the printed specification but is based on  
 CC sequence information supplied to Derwent by the European Patent Office  
 XX  
 SQ Sequence 6301 BP; 1787 A; 59 C; 1247 G; 3207 T; 0 U; 1 Other;  
 Query Match 0.7%; Score 22; DB 6; Length 6301;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3142 TAGTATTTGTTTTTTTGGT 3163  
 DB 3729 TAGTATTTGTTTTTTTGGT 3750  
 RESULT 7  
 ABL3241  
 ID ABL3241 standard; DNA; 9888 BP.  
 XX ABL3241;  
 XX 26-MAR-2002 (first entry)  
 XX Human immune system associated gene SEQ ID NO: 1214.  
 DE Human; immune system disease; cytosine methylation; antiasthmatic;  
 KM antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KM antirheumatic; antirheumatic; antidiabetic; antipsoriatic;  
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
 ds.  
 XX Homo sapiens.  
 XX WO200200928-A2.  
 XX 03-JAN-2002.  
 XX 02-JUL-2001; 2001WO-EP007537.  
 XX 30-JUN-2000; 2000DE-01032529.  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX (EPIC-) EPIDENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K;  
 PI WPI; 2002-130909/17.  
 DR Nucleic acid comprising fragment of chemically modified gene, useful for  
 XX diagnosis and treatment of diseases associated with abnormal cytosine  
 PT methylation.  
 PT Claim 1; SEQ ID NO 1214; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention  
 XX  
 SQ Sequence 9888 BP; 2830 A; 160 C; 2148 G; 4750 T; 0 U; 0 Other;  
 Query Match 0.7%; Score 22; DB 6; Length 9888;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3142 TAGTATTTGTTTTTTTGGT 3163  
 DB 8219 TAGTATTTGTTTTTTTGGT 8240  
 RESULT 8  
 ABL32509  
 ID ABL32509 standard; DNA; 6048 BP.  
 XX ABL32509;  
 XX 26-MAR-2002 (first entry)  
 XX Human immune system associated gene SEQ ID NO: 482.  
 DE Human; immune system disease; cytosine methylation; antiasthmatic;  
 KM antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KM antirheumatic; antirheumatic; antidiabetic; antipsoriatic;  
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
 ds.  
 XX Homo sapiens.  
 XX WO200200928-A2.  
 XX 03-JAN-2002.  
 XX 02-JUL-2001; 2001WO-EP007537.



```

XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 482; 32bp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
CC
SO Sequence 6048 BP; 1296 A; 103 C; 1361 G; 3288 T; 0 U; 0 Other;

Query Match          0.6%; Score 21; DB 6; Length 6048;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 526 TTGGTTTGGTTTGGTTT 546
    |||||
Db 5395 TTGGTTTGGTTTGGTTT 5415

RESULT 9
ABL32031
ID ABL32031 standard; DNA; 6072 BP.
XX
AC ABL32031;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 4.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; anti-anaemic; cytosinatic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for

```

```

PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 4; 32bp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
CC
SO Sequence 6072 BP; 1595 A; 155 C; 1384 G; 2938 T; 0 U; 0 Other;

Query Match          0.6%; Score 21; DB 6; Length 6072;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3145 TATTTGTTTTTTTGGTAA 3165
    |||||
Db 3717 TATTTGTTTTTTTGGTAA 3737

RESULT 10
AAS46448
ID AAS46448 standard; DNA; 8245 BP.
XX
AC AAS46448;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #170.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytosinatic; cancer;
KW tumour; CPG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 170; 27bp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pre-treated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (58) and sequences
CC complementary to (58). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single

```

CC nucleotide polymorphisms and also to be used in an array for analysing  
 CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The  
 CC probes can also be used in a method for ascertaining genetic and/or  
 CC epigenetic parameters for the diagnosis and/or therapy of existing  
 CC diseases or the predisposition to specific diseases, by analysing  
 CC cytosine methylations. The parameters may be compared to another set of  
 CC genetic and/or epigenetic parameters, the differences serving as basis  
 CC for diagnosis and/or prognosis events which are disadvantageous to  
 CC patients. The present sequence is one of the 533 genomic sequences  
 CC derived from tumour suppressor genes and oncogenes. Sequences with even  
 CC numbered Seq ID numbers are the complementary sequence of the  
 CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID  
 CC 535, except for those whose partner sequence is missing). Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 8245 BP; 1978 A; 305 C; 2119 G; 3842 T; 0 U; 1 Other;

Query Match 0.6%; Score 21; DB 4; Length 8245;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3141 TTAGTATTTTGGTTTTTTTG 3161  
 DB 6632 TTAGTATTTTGGTTTTTTTG 6652

RESULT 11  
 ABRN80176  
 ID ABRN80176 standard; DNA; 13453 BP.  
 AC ABRN80176;  
 XX  
 XX  
 DT 15-JUL-2002 (first entry)  
 DE Human chemically modified disease associated gene SEQ ID NO 193.  
 XX  
 XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KM heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KM antidiabetic; cytostatic; anticonvulsant; ds.  
 XX  
 XX Homo sapiens.  
 OS Synthetic.  
 OS  
 XX  
 FN WO200200927-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP007536.  
 XX  
 PR 30-JUN-2000; 2000DE-01032529.  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2002-130908/17.  
 XX  
 PT Novel nucleic acid useful for diagnosis and therapy of diseases  
 PT associated with development genes such as diabetes, comprises a sequence  
 PT of a segment of chemically pretreated DNA of genes associated with  
 PT development.  
 XX  
 XX Claim 1; SEQ ID NO 193; 27pp; English.  
 CC  
 CC The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
 CC of genes associated with development selected from 87 genes listed in the  
 CC specification such as ACCFN, ADFN, or AFDI and comprising one of 350  
 CC sequences (ABN79984-ABN80333) or their complements. The invention is

CC useful for the diagnosis or therapy of diseases associated with  
 CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Currying syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC identity and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (II) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this  
 CC patent did not form part of the printed specification but is based on  
 CC sequence information supplied to Derwent by the European Patent Office

SO Sequence 13453 BP; 2569 A; 384 C; 3712 G; 6788 T; 0 U; 0 Other;

Query Match 0.6%; Score 21; DB 6; Length 13453;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3141 TTAGTATTTTGGTTTTTTTG 3161  
 DB 7909 TTAGTATTTTGGTTTTTTTG 7929

RESULT 12  
 ADA02738/c  
 ID ADA02738 standard; DNA; 52302 BP.  
 AC ADA02738;  
 XX  
 XX  
 DT 06-NOV-2003 (first entry)  
 DE Human CCND2 carcinoma associated gene, SEQ ID NO:1256.  
 XX  
 XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KM prostate; lymphoma; leukemia; cytostatic; gene therapy; drug screening;  
 KM gene; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO2003057146-A2.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 26-DEC-2002; 2002WO-US041414.  
 XX  
 PR 26-DEC-2001; 2001US-00035832.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW;  
 DR WPI; 2003-587068/55.  
 XX  
 PT New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.  
 XX  
 XX Claim 1; SEQ ID NO 1256; 245pp; English.  
 CC  
 CC The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a bioclip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose

CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 8; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2981 GTTCTTTGTTTGGCTTTGG 3001

Db 11586 GTTCTTTGTTTGGCTTTGG 11566

RESULT 13  
 ADB72476/c  
 ID ADB72476 standard; DNA; 52302 BP.

XX ADB72476;

XX 04-DEC-2003 (first entry)

XX Human CCND2 gene.

XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 KM cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Homo sapiens.

XX MO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001MO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 XX cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 304; 2304BP; English.

XX The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2981 GTTCTTTGTTTGGCTTTGG 3001

Db 11586 GTTCTTTGTTTGGCTTTGG 11566

RESULT 14

ADCS5218/c  
 ID ADCS5218 standard; DNA; 52302 BP.

XX ADCS5218;

XX 01-JAN-2004 (first entry)

XX Human CCnd2 genomic sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;  
 KM secreted; transmembrane; intracellular; de.

XX Homo sapiens.

XX MO2003045230-A2.

XX 05-JUN-2003.

XX 02-DEC-2002; 2002MO-US038582.

XX 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of  
 PT the carcinoma-associated (CA) genes, useful for screening for drug  
 PT candidates for diagnosing or treating carcinomas.

XX Claim 1; SEQ ID NO 4; 983BP; English.

XX The invention relates to a recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the fully defined carcinoma-  
 CC associated (CA) genes from the 50 tables given in the specification. The  
 CC CA proteins are secreted, transmembrane or intracellular proteins. The  
 CC recombinant nucleic acids are useful for screening for drug candidates  
 CC for diagnosing or treating carcinomas. Sequences given in ADCS5215-  
 CC ADCS514 represent CA genes of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2981 GTTCTTTGTTTGGCTTTGG 3001

Db 11586 GTTCTTTGTTTGGCTTTGG 11566

RESULT 15

AACS2596/c  
 ID AACS2596 standard; DNA; 420 BP.

XX AACS2596;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 71849.

XX Hybridisation assay; genetic mapping; gene expression control;  
 KM protein identification; signal transduction pathway; metabolic pathway;  
 KM promoter; termination sequence; ss.

[illegible]

PR	29-JUN-1999	99US-0140991P
PR	30-JUN-1999	99US-0141287P
PR	01-JUL-1999	99US-0141424P
PR	01-JUL-1999	99US-0142155P
PR	02-JUL-1999	99US-0142055P
PR	06-JUL-1999	99US-0142390P
PR	08-JUL-1999	99US-0142803P
PR	09-JUL-1999	99US-0142820P
PR	12-JUL-1999	99US-0142977P
PR	13-JUL-1999	99US-0143542P
PR	14-JUL-1999	99US-01436324P
PR	15-JUL-1999	99US-0144005P
PR	16-JUL-1999	99US-0144033P
PR	16-JUL-1999	99US-0144086P
PR	19-JUL-1999	99US-0144335P
PR	19-JUL-1999	99US-0144352P
PR	20-JUL-1999	99US-0144632P
PR	21-JUL-1999	99US-01446814P
PR	21-JUL-1999	99US-0144814P
PR	21-JUL-1999	99US-0145086P
PR	21-JUL-1999	99US-0145088P
PR	22-JUL-1999	99US-0145085P
PR	22-JUL-1999	99US-0145576P
PR	22-JUL-1999	99US-0145813P
PR	22-JUL-1999	99US-0145809P
PR	23-JUL-1999	99US-0145192P
PR	23-JUL-1999	99US-0145145P
PR	23-JUL-1999	99US-0145218P
PR	23-JUL-1999	99US-0145524P
PR	26-JUL-1999	99US-0145766P
PR	27-JUL-1999	99US-0145813P
PR	27-JUL-1999	99US-0145818P
PR	28-JUL-1999	99US-0145911P
PR	28-JUL-1999	99US-01459512P
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146388P
PR	03-AUG-1999	99US-0146389P
PR	03-AUG-1999	99US-0147038P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147302P
PR	05-AUG-1999	99US-0147192P
PR	06-AUG-1999	99US-0147260P
PR	06-AUG-1999	99US-0147303P
PR	06-AUG-1999	99US-0147416P
PR	09-AUG-1999	99US-0147493P
PR	09-AUG-1999	99US-0147935P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148311P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149902P
PR	23-AUG-1999	99US-0149930P
PR	23-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	30-AUG-1999	99US-0151080P
PR	31-AUG-1999	99US-0151303P
PR	01-SEP-1999	99US-0151930P
PR	07-SEP-1999	99US-0152363P
PR	10-SEP-1999	99US-0153070P

PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query March 0.6%; Score 20; DB 3; Length 420;  
Best Local Similarity 100.0%; Pred.No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACAACAAACAAATTAG 21  
|||||  
Db 378 ACAACAAACAAATTAG 359

Search completed: March 30, 2004, 09:26:44  
Job time : 1138.92 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 03:15:12 ; Search time 12253.4 Seconds

(without alignments)  
11584.370 Million cell updates/sec

Title: US-10-030-829-1

Perfect score: 3275

Sequence: 1 gacaaacaacaacaataa.....caaatatatacggtacaaat 3275

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_ov:\*  
22: em\_or:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_pi:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_mus:\*  
34: em\_hgt\_pln:\*  
35: em\_hgt\_rod:\*  
36: em\_hgt\_man:\*  
37: em\_hgt\_vrt:\*  
38: em\_sy:\*  
39: em\_hgt\_hum:\*  
40: em\_hgt\_mus:\*  
41: em\_hgt\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3275	100.0	3275	6	AX078760	AX078760 Sequence
2	3275	100.0	81365	8	AB025633	AB025633 Arabidops
3	2254	68.8	2254	8	AF239719	AF239719 Arabidops
4	972	29.7	2162	8	BT002944	BT002944 Arabidops
5	963	29.4	1878	6	AX078761	AX078761 Sequence
6	963	29.4	1909	8	BT004380	BT004380 Arabidops
7	541	16.5	650	8	ATHS28171	ATHS28171 Arabidops
8	45	1.4	45	8	ATHS27741	ATHS27741 Arabidops
9	24	0.7	170765	2	AC102284	AC102284 Mus muscu
10	23	0.7	554	3	AF057056	AF057056 Uroleucon
11	23	0.7	101080	2	BX004860	BX004860 Dario rer
12	23	0.7	101241	5	AL845510	AL845510 Zebrafish
13	23	0.7	126716	9	AC107028	AC107028 Homo sapi
14	23	0.7	152345	5	BX005128	BX005128 Zebrafish
15	23	0.7	180925	2	AC048381	AC048381 Homo sapi
16	22	0.7	6136	6	AX344692	AX344692 Sequence
17	22	0.7	2100	3	D84313	D84313 Drosophila
18	22	0.7	6301	6	AX344692	AX344692 Sequence
19	22	0.7	9888	6	AX346143	AX346143 Sequence
20	22	0.7	80318	5	AL645798	AL645798 Zebrafish
21	22	0.7	81835	8	ATFC9	ATFC9 Arabidops
22	22	0.7	110000	2	AC107170_1	AC107170_1 Arabidops
23	22	0.7	133120	9	HS393P23	HS393P23 Continuation (2 of
24	22	0.7	141509	9	AC006044	AC006044 Homo sapi
25	22	0.7	150030	2	AC145578	AC145578 Mus muscu
26	22	0.7	154923	9	AC026185	AC026185 Homo sapi
27	22	0.7	158661	10	AC074229	AC074229 Mus muscu
28	22	0.7	174612	9	AC018494	AC018494 Homo sapi
29	22	0.7	174641	2	AC016346	AC016346 Homo sapi
30	22	0.7	175569	10	AC109214	AC109214 Mus muscu
31	22	0.7	177529	2	AC119271	AC119271 Homo sapi
32	22	0.7	177834	9	AC018811	AC018811 Homo sapi
33	22	0.7	178928	2	AC121725	AC121725 Rattus no
34	22	0.7	179134	2	AC119607	AC119607 Rattus no
35	22	0.7	181166	9	AC087897	AC087897 Homo sapi
36	22	0.7	182509	9	AC112211	AC112211 Homo sapi
37	22	0.7	183936	10	AC007978	AC007978 Mus muscu
38	22	0.7	191391	2	AC128938	AC128938 Rattus no
39	22	0.7	192173	9	AC107462	AC107462 Homo sapi
40	22	0.7	193523	5	BX001014	BX001014 Zebrafish
41	22	0.7	197909	10	AL928719	AL928719 Mouse DNA
42	22	0.7	198067	8	ATCHRIV47	ATCHRIV47 Arabidops
43	22	0.7	201208	2	AC107451	AC107451 Mus muscu
44	22	0.7	201803	2	AC113572	AC113572 Canis fam
45	22	0.7	205209	2	AC137942	AC137942 Mus muscu

# ALIGNMENTS

RESULT 1	AX078760	3275 bp	DNA	linear	PAT 22-FEB-2001
LOCUS	AX078760				
DEFINITION	Sequence 1 from Patent WO0105951.				
ACCESSION	AX078760				
VERSION	AX078760.1	GI:13158379			
KEYWORDS					
SOURCE					
ORGANISM	Arabidopsis thaliana (thale cress)				
REFERENCE 1	Becilj, C., Elmayer, T. and Vaucheret, H.				
AUTHORS	Novel gsg3 plant gene and use thereof				
TITLE					

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0105951-A 1 25-JAN-2001;  
 AVENTIS CHROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE  
 AGRONOMIQUE (FR)  
 FEATURES Location/Qualifiers  
 source 1..3275  
 /organism="Arabidopsis thaliana"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3702"  
 primer\_bind 693..715  
 /note="P356AD"  
 primer\_bind complement(2926..2952)  
 /note="P356Y"  
 ORIGIN  
 Query Match 100.0%; Score 3275; DB 6; Length 3275;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAACAAACAAAATTAGCAAGTCATGTTCCGACATTAATTAATAGGGAAACA 60  
 Db 1 GACAAACAAACAAAATTAGCAAGTCATGTTCCGACATTAATTAATAGGGAAACA 60

QY 61 TTAAGTTAAGCGAAAAAGAAAAAAGGTACAAAAATGAAAACAAAATCAATGAT 120  
 Db 61 TTAAGTTAAGCGAAAAAAGAAAAAAGGTACAAAAATGAAAACAAAATCAATGAT 120

QY 121 GAAAAATTGAGTCCAGAAATGGAAGAAACGAGCCGTTTAGACTTAATTAAGCTTCTC 180  
 Db 121 GAAAAATTGAGTCCAGAAATGGAAGAAACGAGCCGTTTAGACTTAATTAAGCTTCTC 180

QY 181 ATTGTCTCTCTTCGTCAGATTATTTTCTCTCCGAGCTCTGACTCACTAC 240  
 Db 181 ATTGTCTCTTCGTCAGATTATTTTCTCTCCGAGCTCTGACTCACTAC 240

QY 241 TCTCCGCGCTTTAACTTACGTTCTCCGTCCTTACTCTGTAAGTTTCTGCTTAAG 300  
 Db 241 TCTCCGCGCTTTAACTTACGTTCTCCGTCCTTACTCTGTAAGTTTCTGCTTAAG 300

QY 301 CCTCGATCGGCTCACCGCATGCACTGTCGTCGATTTCTCTTTCTTCGCTGAAAA 360  
 Db 301 CCTCGATCGGCTCACCGCATGCACTGTCGTCGATTTCTCTTTCTTCGCTGAAAA 360

QY 361 ATTGCCCTAATGTTCTGATTTGCAAGGTTTTGTCATGAGTTTACTTTTCCCTATA 420  
 Db 361 ATTGCCCTAATGTTCTGATTTGCAAGGTTTTGTCATGAGTTTACTTTTCCCTATA 420

QY 421 TTTTAAAGTCTTAGGTAAGCAATCCGCGCTTACTGTTTGTTCATTTTGTGCT 480  
 Db 421 TTTTAAAGTCTTAGGTAAGCAATCCGCGCTTACTGTTTGTTCATTTTGTGCT 480

QY 481 TTCAACGTTTAGTGCATGATCGAGATTGACTGTGAAAAATCCTCGTTTTTGGTT 540  
 Db 481 TTCAACGTTTAGTGCATGATCGAGATTGACTGTGAAAAATCCTCGTTTTTGGTT 540

QY 541 TTGTTCAATAAATCGAATTGATCTACCTTTTGTCCTTGATGTTGTTTTGAGCCT 600  
 Db 541 TTGTTCAATAAATCGAATTGATCTACCTTTTGTCCTTGATGTTGTTTTGAGCCT 600

QY 601 ATGGCTGTGGCTTTGTAATTAATCTCACTGATGTGATTTGATTTGGTAAGT 660  
 Db 601 ATGGCTGTGGCTTTGTAATTAATCTCACTGATGTGATTTGATTTGGTAAGT 660

QY 661 ACTGTGGGTTTCTTTGGTGCATAGGTTGTAATAATGAGTTCTAGGGCTGATCAATG 720  
 Db 661 ACTGTGGGTTTCTTTGGTGCATAGGTTGTAATAATGAGTTCTAGGGCTGATCAATG 720

QY 721 CTAAAGAAAAAGACGTTCAAGGTGTTAATAGGCTGAGTTGAACAGTTGTTCAAGTT 780  
 Db 721 CTAAAGAAAAAGACGTTCAAGGTGTTAATAGGCTGAGTTGAACAGTTGTTCAAGTT 780

QY 781 TGGCAGGAGAGAGACGTTGCTTCAACAATGATGAGAGAGAGTGGAGAGTCAATTTCCA 840  
 Db 781 TGGCAGGAGAGAGACGTTGCTTCAACAATGATGAGAGAGAGTGGAGAGTCAATTTCCA 840

QY 841 AGAAGAACAAAGAACAAACGAGAAACATTTGAAAAAATTGGGTTTCTCAGAAATTCGA 900  
 Db 841 AGAAGAACAAAGAACAAACGAGAAACATTTGAAAAAATTGGGTTTCTCAGAAATTCGA 900

QY 901 ATCTCTCTAGAGCTTGGGGTGTGTCACAGTAAGGAGAGAGTATGCAATCTTGGAGAG 960  
 Db 901 ATCTCTCTAGAGCTTGGGGTGTGTCACAGTAAGGAGAGAGTATGCAATCTTGGAGAG 960

QY 961 GAAACATATATCCGGAGAGAGTAAAGCAATGTTGGGGCATTCAAGCTTAATATATCTG 1020  
 Db 961 GAAACATATATCCGGAGAGAGTAAAGCAATGTTGGGGCATTCAAGCTTAATATATCTG 1020

QY 1021 GTGGGGAGCAGGCTTGAAGAGAAATGATTAACAATTTGTGCAACCCCACTGTAT 1080  
 Db 1021 GTGGGGAGCAGGCTTGAAGAGAAATGATTAACAATTTGTGCAACCCCACTGTAT 1080

QY 1081 CTGCGCCCTCTTTGGAAGAGAGTGAATTTGGCAAGAGAGGTTCTGCTCAGACACA 1140  
 Db 1081 CTGCGCCCTCTTTGGAAGAGAGTGAATTTGGCAAGAGAGGTTCTGCTCAGACACA 1140

QY 1141 CAGCTGTCAGAGATTCTCTGACGTGAGAGATGATGATATATGCTTCTGAGAGAGAGA 1200  
 Db 1141 CAGCTGTCAGAGATTCTCTGACGTGAGAGATGATGATATATGCTTCTGAGAGAGAGA 1200

QY 1201 ATGATTTCCGATGCTTTGATGATGATTTCTGATGACGACTTGCAAGTATATGATCTCG 1260  
 Db 1201 ATGATTTCCGATGCTTTGATGATGATTTCTGATGACGACTTGCAAGTATATGATCTCG 1260

QY 1261 ATGTAGTCAAAAGAGCAGATGATCAAGAAAGAGAAATGAGGTTCAAAAAGTTCTTG 1320  
 Db 1261 ATGTAGTCAAAAGAGCAGATGATCAAGAAAGAGAAATGAGGTTCAAAAAGTTCTTG 1320

QY 1321 GCAGCTTGATAGCTTGTGATGAGCAGATTAATGAACACAGAGGAGTGCGAATGTC 1380  
 Db 1321 GCAGCTTGATAGCTTGTGATGAGCAGATTAATGAACACAGAGGAGTGCGAATGTC 1380

QY 1381 CAGCTTGTCAGAACGAGCCTGCTGTCATGATTTGGTAATACCTGCAACCTCTACTAGCTC 1440  
 Db 1381 CAGCTTGTCAGAACGAGCCTGCTGTCATGATTTGGTAATACCTGCAACCTCTACTAGCTC 1440

QY 1441 ATGCGAGAGCAAAAAGAGCTAGGCGAGTTAACTCCATGAGAAATTTGGCTGAAGTTTAA 1500  
 Db 1441 ATGCGAGAGCAAAAAGAGCTAGGCGAGTTAACTCCATGAGAAATTTGGCTGAAGTTTAA 1500

QY 1501 AAAAGATCTAAGATGAGAGGCGATCTGTATTCCTTTGTGTAAGATTATGAGCACT 1560  
 Db 1501 AAAAGATCTAAGATGAGAGGCGATCTGTATTCCTTTGTGTAAGATTATGAGCACT 1560

QY 1561 GGAAGGTTTGGGTGAGAGTGAAGAAAGATTAAGAAATGTCGAGCTCCAAATGTCATCA 1620  
 Db 1561 GGAAGGTTTGGGTGAGAGTGAAGAAAGATTAAGAAATGTCGAGCTCCAAATGTCATCA 1620

QY 1621 TCATGAATCTAGACTGATGATAGAGAGATTAAGAGTGAATTTCTCTCTTTTAC 1680  
 Db 1621 TCATGAATCTAGACTGATGATAGAGAGATTAAGAGTGAATTTCTCTCTTTTAC 1680

QY 1681 TTCTTAATTTTCTCTTGCATTTCTAGATTTTGAATTTGAATTTGATGAGCTCGAG 1740  
 Db 1681 TTCTTAATTTTCTCTTGCATTTCTAGATTTTGAATTTGAATTTGATGAGCTCGAG 1740

QY 1741 ATGGGCAACCAAGAGCTGCTGGAATTAATCTTCAAGTATGAGGCTTAAAGAGCAGCAT 1800  
 Db 1741 ATGGGCAACCAAGAGCTGCTGGAATTAATCTTCAAGTATGAGGCTTAAAGAGCAGCAT 1800

QY 1801 TCTATAGTCAACAGGCGCATGTCGAGATGATGTTCTGATGTTTGAAGAGCAGTCCACT 1860  
 Db 1801 TCTATAGTCAACAGGCGCATGTCGAGATGATGTTCTGATGTTTGAAGAGCAGTCCACT 1860

QY 1861 GGTATTTTGAAGCCGAAACGCTTCCACCGGAGTTAGCTGAGATGGGTTGATGAATTT 1920  
 Db 1861 GGTATTTTGAAGCCGAAACGCTTCCACCGGAGTTAGCTGAGATGGGTTGATGAATTT 1920



QY 1921 GCCTGGGAGTGAAGCCAGATGTTTCTGAGAGTGTCCCAACTGATGCTCTT 1980  
 DB 1921 GCCTGGGAGTGAAGCCAGATGTTTCTGAGAGTGTCCCAACTGATGCTCTT 1980  
 QY 1981 GCAAGCAAGCAAGTCTGAGCAATTCATCAACACTCTCAAGGTTCTTCCCCAAGA 2040  
 DB 1981 GCAAGCAAGCAAGTCTGAGCAATTCATCAACACTCTCAAGGTTCTTCCCCAAGA 2040  
 QY 2041 AATTGATATATAGCTTTTGAAGTGTGATGGAATTTAAAGTTTGTGGTCCGTGTA 2100  
 DB 2041 AATTGATATATAGCTTTTGAAGTGTGATGGAATTTAAAGTTTGTGGTCCGTGTA 2100  
 QY 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAAACAAGCTGAATTTGAGTTG 2160  
 DB 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAAACAAGCTGAATTTGAGTTG 2160  
 QY 2161 AATCATATCCAGAGATGTTGTAAGAGCTGAGGAGATCTCTGAGGACAATCAGCAG 2220  
 DB 2161 AATCATATCCAGAGATGTTGTAAGAGCTGAGGAGATCTCTGAGGACAATCAGCAG 2220  
 QY 2221 CTGAACTACTTTAAGAACAGCTCTCAAAAACAGAACAGCAAGCTGCTTGAAGAA 2280  
 DB 2221 CTGAACTACTTTAAGAACAGCTCTCAAAAACAGAACAGCAAGCTGCTTGAAGAA 2280  
 QY 2281 TCTCTGGAATTTATGAGCCAGAAAGCTGCTGTAAGTCTGAGAGATTAATCGATGCTGAGA 2340  
 DB 2281 TCTCTGGAATTTATGAGCCAGAAAGCTGCTGTAAGTCTGAGAGATTAATCGATGCTGAGA 2340  
 QY 2341 CAGAGAACTAAGATGAGCATGAAACAGAACAGGAGAGATGATGATTTTCTTGAAGAA 2400  
 DB 2341 CAGAGAACTAAGATGAGCATGAAACAGAACAGGAGAGATGATGATTTTCTTGAAGAA 2400  
 QY 2401 CACAAATTTGACATTTTGTATTAATCTGATTCATCAATTTTGTATATTTGTCCAA 2460  
 DB 2401 CACAAATTTGACATTTTGTATTAATCTGATTCATCAATTTTGTATATTTGTCCAA 2460  
 QY 2461 AAAACCTGTGTGTTGAGATGATGATGACAGACAGAGTTTTCATGAGATTCATCA 2520  
 DB 2461 AAAACCTGTGTGTTGAGATGATGATGACAGACAGAGTTTTCATGAGATTCATCA 2520  
 QY 2521 CAGATCCATGAAAGAGAGAGAGAGAGAGATTTTGAAGTGTGAGCAGCAGAGAA 2580  
 DB 2521 CAGATCCATGAAAGAGAGAGAGAGAGAGATTTTGAAGTGTGAGCAGCAGAGAA 2580  
 QY 2581 CGTCCCAAGTGTGTGCGCAGACAGACAGACATTTATCCCTCTACCAATGACGATTC 2640  
 DB 2581 CGTCCCAAGTGTGTGCGCAGACAGACAGACATTTATCCCTCTACCAATGACGATTC 2640  
 QY 2641 CGAAAGAGTATATGTAATCAATTAATCCCTGCGGCTTTTGTGTTTCAAACTTA 2700  
 DB 2641 CGAAAGAGTATATGTAATCAATTAATCCCTGCGGCTTTTGTGTTTCAAACTTA 2700  
 QY 2701 AGAGTAACTGAATTTCCGCGTTTGAATTTCTTTGCGAGCTGAGAAAGTGTCAAGCTTC 2760  
 DB 2701 AGAGTAACTGAATTTCCGCGTTTGAATTTCTTTGCGAGCTGAGAAAGTGTCAAGCTTC 2760  
 QY 2761 ATGAGCTTTCAAGAAAGAGTGAAGAGATTTGTGAAGAGAGGAGATGCTGATAAA 2820  
 DB 2761 ATGAGCTTTCAAGAAAGAGTGAAGAGATTTGTGAAGAGAGGAGATGCTGATAAA 2820  
 QY 2821 GATCAAGAGAGAGATGAGAGACATGAGAGAGGAGATCAAGAGAGATTTGATCTG 2880  
 DB 2821 GATCAAGAGAGAGATGAGAGACATGAGAGAGGAGATCAAGAGAGATTTGATCTG 2880  
 QY 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTAACAAGCATGCTTCAAAATGAA 2940  
 DB 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTAACAAGCATGCTTCAAAATGAA 2940  
 QY 2941 GATGATTTGAGACAAAAGTCTGATGACAAAGCAAGATTAAGTTTGTGCTTTTG 3000  
 DB 2941 GATGATTTGAGACAAAAGTCTGATGACAAAGCAAGATTAAGTTTGTGCTTTTG 3000  
 QY 3001 GTATGTGGAAGAGATGAGAGATCTGAGAGATCCATTTAAATATAGAGCAAAATCTAAGAA 3060

DB 3001 GTATGTGGAAGAGATGAGAGATCTGAGAGATCCATTTAAATATAGAGCAAAATCTAAGAA 3060  
 QY 3061 GATTATGATTTATTTATCTCCATTTTATAGTACGAGATCTTAAGAGCATTAAGTTCTT 3120  
 DB 3061 GATTATGATTTATTTATCTCCATTTTATAGTACGAGATCTTAAGAGCATTAAGTTCTT 3120  
 QY 3121 GTGACTAAACCAAGTTCCTTAGTATTTGTTTGTGTAATTTCAATGTAAG 3180  
 DB 3121 GTGACTAAACCAAGTTCCTTAGTATTTGTTTGTGTAATTTCAATGTAAG 3180  
 QY 3181 TTAGACATATTACCAACGTCAGAGTGAATCAAGATGCAAAATCAATGTTT 3240  
 DB 3181 TTAGACATATTACCAACGTCAGAGTGAATCAAGATGCAAAATCAATGTTT 3240  
 QY 3241 AGAATTTATATCTACCAAAATATATGCTCAAT 3275  
 DB 3241 AGAATTTATATCTACCAAAATATATGCTCAAAAT 3275

RESULT 2  
 AB025633  
 LOCUS  
 DEFINITION  
 Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MQM1.  
 ACCESSION  
 AB025633  
 VERSION  
 AB025633.2  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Burkholderia; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases)  
 Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E., Kotani, H.  
 and Tabata, S.  
 TITLE  
 Structural analysis of Arabidopsis thaliana chromosome 5. X.  
 Sequence features of the regions of 3,076,755 bp covered by sixty  
 pl and TAC clones  
 JOURNAL  
 DNA Res. 7 (1), 31-63 (2000)  
 MEDLINE  
 20181125  
 PUBMED  
 10718197  
 REFERENCE  
 2 (bases 1 to 81365)  
 AUTHORS  
 Nakamura, Y.  
 TITLE  
 Direct Submission  
 Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research  
 Institute, Department of Plant Gene Research; 1532-3, Yana,  
 Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,  
 Tel:81-438-52-3935, Fax:81-438-52-3934)  
 On Sep 15, 2000 this sequence version replaced gi:4589439.  
 COMMENT  
 Address for correspondence: kaos@kazusa.or.jp  
 For the latest information on annotation of this clone, please see  
 http://www.kazusa.or.jp/kaos/cgi-bin/sgd\_graph.cgi?c=MQM1  
 Genes with similarity to proteins in the databases are described in  
 'product' or 'note' qualifiers. Genes that have no significant  
 protein similarity are described as 'unknown protein'.  
 The software programs used to predict genes include: Grail  
 (Informatics Group, Oak Ridge National Laboratory,  
 http://compbio.ornl.gov/Grail-1.3/),  
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),  
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of  
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and  
 SplicePredictor (Volker Brendel, Stanford University,  
 http://gremlin.zool.laastate.edu/cgi-bin/sp.cgi).  
 Genes encoding tRNAs are predicted by tRNAscan-SE  
 (Sean Eddy, Washington University School of Medicine, St. Louis,  
 http://genome.wustl.edu/eddy/tRNAscan-SE/).  
 This sequence may not be the entire insert of this clone. It may be  
 shorter because we remove overlaps between neighboring submissions.  
 The 5' clone is K19M13 and the 3' clone is M2011.  
 Location/Qualifiers  
 1..81365  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"

FEATURES

source

/strain="Colombia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 /clone="MOM1"  
 /clone.lib="Miteui P1"  
 join(3184, .4403,3474, .3595,3681, .3723,3777, .3898, 4042, .4198,4445, .4515,4601, .4714,4888, .4988,5300, .5420)  
 /note="gb|AA01580.1  
 gene\_id:MOM1.24  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BA097251.1"  
 /db\_xref="GI:8809710"  
 /translation="MHWQGTSTVDESDPAHKYDDETPPKHYQLEPISTKVNDADQNS  
 YSINGTSSNFWVCHLEDGVNYIDESSRERTSNAGKQNTLVADVDYLA1EV  
 OPLDDNKNIEKTAFGKYRDLTVLADVHTLLFSADCFVFFMFAIICYFLISGG  
 ACPDPMQNGIKETLTHGASYKVLAVGVIMWATLISIKEGYSICIDLTAE  
 KRSATTVIPGQPAELVIGDSGSEHSLRADNSPDLIGSDEIVLRLFIKRW  
 FFKANSFPHRSCSKRSQNL"  
 join(6317, .6428,6508, .6629,6714, .6756,6841, .6932, 7121, .7169,7396, .7466,7550, .7663,7785, .7862,8112, .8181)  
 /note="gb|AA01580.1  
 gene\_id:MOM1.23  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BA097250.1"  
 /db\_xref="GI:8809709"  
 /translation="MQGTSTVESPALNQADETPISKTVDNTKNIFFYDGDGDAFL  
 PALENIQISEPYPGHELOACGYSINGTSCNFWCHLEDSSVNYIDAKKPNYLV  
 ADVGCLAIIEVPLDDNRKGEVLVFPANDRKILCHPEMOSINDIKTLHASYK  
 SLAIGVHLEATLISIREGTYIKCNNDLTTEKFSATAVYIPREKPAEVIIGSD  
 GSCHCARVNEPMDIRLCKEKKYRLVLYVYVEBOKL"  
 complement(join(8642, .8705,8794, .8863,8978, .9387, 9584, .10347))  
 /note="gene\_id:MOM1.22  
 unknown protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BA097249.1"  
 /db\_xref="GI:8809708"  
 /translation="MSWNRKSSGWTAPDLKOROKGLESEVEDPFPVSTVNASF  
 GVGRFRRHREPEKSSVLLPPSRFPALTEKDCNGERGCGCKRKPDTLSLPNS  
 HDIAFLKEMSMADNDLIRVLTSTEDFEMALAFLKMGSGKDEDEPSIKLYG  
 SSSNRSEKTEFEKTVTSVYKMAASRTEPDAGKYLDNSGSSFLVNAADNEKFPDI  
 SELDSIIRLOSLPIREWEEDDLYSHRKDLKVRASLHNSRAQNAFORYDASA  
 KOSDKARNDLAELKINAELAKKTIIGHKNDIKWLSLHGAHEAVALOERLQW  
 IBEHFTVNSVSPNRGRSKALRSASQEPFGHLDSEGMHCQRTSRELRNLQVITG  
 IGHSGQASLPAAKTFEFDNRYRDETRPGVITVRPFRHS"  
 complement(12417, .13424)  
 /note="contains similarity to unknown protein  
 gb|AA27018.1  
 gene\_id:MOM1.21  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BA097248.1"  
 /db\_xref="GI:8809707"  
 /translation="MATDSOPNOKTLPLKTRIALTVISTMDNORPDGTINRRFLR  
 LPFRAPPNPKPVNISTDEVDOSRDLFRITLPHVSGDKIPVVPFGGFEALS  
 NMDSPVDNRRFARKLPAVVISVNRILPEHHYRAGSDGALPKYIENNGSLIAP  
 NADISCFPAGDSAGNIAHNVARIICRRERSFTYKYLIGLISITOPFFGSERTAE  
 KOLVAGPLVSPDITMCKAMGNDRHNAVNGEPNAVISGLDIPETVNAVAGFPL  
 KMQRYSYELKICGKATLIEYPMHAFYIFPELPEAGQLIMRIKDVDRVASLS  
 A"  
 join(4922, .14981,15072, .15167,15563, .15666,15755, .15891, 16096, .16178)  
 /note="gene\_id:MOM1.20"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="50S ribosomal protein L24"  
 /protein\_id="BA097247.1"

/db\_xref="GI:8809706"  
 /translation="MGKAAKRLIRKMLIGDNVMIIRGDKSETGIRKVISQNR  
 VIVSGKILKIKHKGPDHGGIFVEKPLHSAVQVVDVTVGRCKGVKLYDEGTR  
 VRVARGTGTSGSIIIPREILIRATPREPTLGPDTPEFWEQTYDAKTKGMPDL"  
 join(16823, .16952,17143, .17320,17412, .17493,17675, .17851, 17943, .18110,18199, .18390)  
 /note="gene\_id:MOM1.19"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="26S proteasome, non-ATPase regulatory subunit"  
 /protein\_id="BA097246.1"  
 /db\_xref="GI:8809705"  
 /translation="MERLQRIFGAGGGLGHASPDSPDLTSEQVYSSIALIKMLKHG  
 RAGVMEVGMILGEFVEDEYTVAGVDVAMQSGVSVVEAVDFQTNMDMLKOTG  
 RPEWYVGHSHRPGCKLGGVDINTQSSPEALNOAAVAVVDVIAQYKGVDAFR  
 SINPOTMLGEPRTTSLNHLKPKSLQALIHGLNRYVIAINRYRNELEBKMLN  
 LHKRWDTGLTLRFDTHSKNEQTVQDMSLAKYNRAVVEBELSEKALIVNVR  
 QDAKKHLEHNSNMSSNIYQTLGMDLVTF"  
 complement(join(18965, .19043,19240, .19298,19419, .19536, 19666, .19739,19908, .19961,20387, .20530))  
 /note="contains similarity to unknown protein  
 emb|CAA2897.1"  
 gene\_id:MOM1.18  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BA097245.1"  
 /db\_xref="GI:8809704"  
 /translation="MDKKNQALIEKKMLVGMVEVEQOQADESSISFEDNRICAL  
 TTKORFVGATCISAGLTCILSMVFNPNPKRGTITFLGMLAGSTAPLIGPROV  
 TMDLDPATVATLYALIASIIITLFCALYVRKLLTLALITLFFGLIYSSLYIFEAR  
 TVMSKIEWTCFDTBE"  
 join(12672, .23634,23708, .23999,24111, .24355,24458, .24624, 24715, .24925)  
 /note="emb|CAB62356.1  
 gene\_id:MOM1.17  
 similar to unknown protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BA097244.1"  
 /db\_xref="GI:8809703"  
 /translation="MSRAGPMSEKKNVQGGYRPEVEQVQGLAGTRILASODDCEM  
 EVISKNNKRGNTSGKTWVSQNSNPAPMGQOQGRSNTVSGGANNVSGKNGRG  
 IOANISGGRALSKRTDNNYPAPPVSPPELEGNNARQSGAHOHTAVQFPVEDD  
 VDNASEBNDSDALDDSDDLASDSDYSDQKSHGSKQNMKEKFPSSGDSLSIBQ  
 INEPORQNHCAQNGAPALIDWYNILDAVTRKAGARVRLHFAVLKDLQMRG  
 ASVPIICGIVQCGHGDDEYELVEMPAVITINTRLDKDNKMLQMGQELIEY  
 DVEYLRARHYSVQGGHGMGVMESSATVYIELERHDLAEGLDRIIMQGRSM  
 FSGVGRKOLYGLATKODDIDHNSQGTTRKFLKYSQWVWVXELRQISDNQOLNY  
 FKNKLSKNNKATYLESLTESEKTRPALTANDIVRQRTQWQHONSEMDADHFRF  
 MDSIQIHERDPAENRPMIQQGRNAVVOQQQNTINPSSNDCKRAEVSSTIEF  
 QKEKHEEFVEEREMILIKQEKKEMDKRHHBELFDEKEFDEALBQMLYRGHANE  
 D"  
 join(25614, .25838,26061, .26207,26324, .26438,26516, .26610, 26710, .26813,27522, .27765,27943, .28036,28115, .28305, 28385, .28604,28688, .28855,28944, .29092)  
 /note="gb|AAC97420.1  
 gene\_id:MOM1.16  
 strong similarity to unknown protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BA097243.1"  
 /db\_xref="GI:8809702"  
 /translation="NAPPAQGTAAVAAAGADGAPPOOGRGSGTIGVRIAVFWY  
 ASKFSRQKPMNDPAEQLMTNINLFHKESSLDMPFYLSEQEKDPRGDRALYHETN  
 ILYAVWTESIRTSLLTYPSBETLQNNSLYAHIFASGFPIDTDEYOLNSFSR  
 THAVATYPPKQKNNKKSLGSPDOSDESEBEVEVQDKSDPEKEVEVEMISLKN  
 VTINLVDFOTFADILAVETPTGQNYPTIYFNEFMLIRKQIPVNETVEPLNLEISF  
 ISMKMOLFQOVDOSFQORQYSGMLDSEDELMKFTFLGEGNLYLLGIMFMSHVSF  
 DELAFKNDIOFMNNKSMWGLSASVYNAFICQAFIYFLILNDNSMILLASGVCV  
 IEFKTIQKAMRIYDRSGMIRLPHDEESAYKTYEDILAIKFLSVYLLIYVIG  
 STYSLAERHKSMTYSLSTSCVYMGFTLMCEQLFTINTKLSVAHLPRQMTYK  
 LNTIITDIFAFYIKPILIRLSVFREDVIFILYLYQRVVYVPDKTRVNEFGGGEDET  
 AEKTIITEEDDKKTN"

CDS		complement(join(29439..29576,29661..29885,29985..30341..30493,30579..30722,30956..31484))		/note="gene_id:MGM1.15"	
Query Match		100.0%;	Score 3275;	DB 8;	Length 81365;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 3275;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GACAAACAAACAAATTTAGCAGTATGTTGCTAGCAATTAATTAATAGTGGAA	CA	60	
DB	21977	GACAAACAAACAAATTTAGCAGTATGTTGCTAGCAATTAATTAATAGTGGAA	CA	22036	
QY	61	TTAAGTTAAGCGAAAAAGAAAAAGGTACAAAAATGAAAAATCAAACTGAAT		120	
DB	22037	TTAAGTTAAGCGAAAAAGAAAAAGGTACAAAAATGAAAAATCAAACTGAAT		22096	
QY	121	GAAAAATTGAGTCCAGAAATCGAAAAACGAGCCGTTTAAAGCTTAATAGCTTC		180	
DB	22097	GAAAAATTGAGTCCAGAAATCGAAAAACGAGCCGTTTAAAGCTTAATAGCTTC		22156	
QY	181	ATTGCTCTTCTTCTGCAATTTATTTCTCCCTCCGAGCTCAGTCACTACTCTC		240	
DB	22157	ATTGCTCTTCTTCTGCAATTTATTTCTCCCTCCGAGCTCAGTCACTACTCTC		22216	
QY	241	TCTCCGCGCTTTAACTTAAGTCTCCGCTTCTAGCTCTAGATTTCGCTTAGAG		300	
DB	22217	TCTCCGCGCTTTAACTTAAGTCTCCGCTTCTAGCTCTAGATTTCGCTTAGAG		22276	
QY	301	CTTCCGATCGCTCAGCGCATGCTCTGTCGATTTCTCTTTTCTTCCGTGAAAA		360	
DB	22277	CTTCCGATCGCTCAGCGCATGCTCTGTCGATTTCTCTTTTCTTCCGTGAAAA		22336	
QY	361	ATTGCCCTAATGTTCTGATTTGCAAGTTTGTGCTAGAGGTTACTTTTCCCAT		420	
DB	22337	ATTGCCCTAATGTTCTGATTTGCAAGTTTGTGCTAGAGGTTACTTTTCCCAT		22396	
QY	421	TTTAAATAGTCTTAAGTAAACGATACCTGCGCTTACTGTTTGTCTATTTGTG		480	
DB	22397	TTTAAATAGTCTTAAGTAAACGATACCTGCGCTTACTGTTTGTCTATTTGTG		22456	
QY	481	TTCAACGTTTAACTGCTGATCGAGATTTGACTGTGAAAAATCTTTCGTTTGT		540	
DB	22457	TTCAACGTTTAACTGCTGATCGAGATTTGACTGTGAAAAATCTTTCGTTTGT		22516	
QY	541	TTGTTTCAATTAATCGATTTGATCTACTTTGTGCTTGAATTTGTTTGAACCT		600	
DB	22517	TTGTTTCAATTAATCGATTTGATCTACTTTGTGCTTGAATTTGTTTGAACCT		22576	
QY	601	ATGCGTTGTTGCTGTTTAACTTCACTGATGATGATTTGAGATTGTTGTTAGT		660	
DB	22577	ATGCGTTGTTGCTGTTTAACTTCACTGATGATGATTTGAGATTGTTGTTAGT		22636	
QY	661	ACTGTGGGTTTCTTGTGCTATAGTTGTAATAATGAGTTCTAGGGCTGTGCAAT		720	
DB	22637	ACTGTGGGTTTCTTGTGCTATAGTTGTAATAATGAGTTCTAGGGCTGTGCAAT		22696	
QY	721	CTAAGAAAAAGACGTTCAAGGTGTGTTAAGGCTTGAACAGTTGTTCAAGTT		780	
DB	22697	CTAAGAAAAAGACGTTCAAGGTGTGTTAAGGCTTGAACAGTTGTTCAAGTT		22756	
QY	781	TGCGAGGAGCGAGCTGGCTTCTCACAAATGATGAGAGAGTGGAGGTCAATTT		840	
DB	22757	TGCGAGGAGCGAGCTGGCTTCTCACAAATGATGAGAGAGTGGAGGTCAATTT		22816	
QY	841	AGAAAGCAAGAACAAACCAAGAAACATTTGAGAAAACTTGGGTTTCTCAGAA		900	
DB	22817	AGAAAGCAAGAACAAACCAAGAAACATTTGAGAAAACTTGGGTTTCTCAGAA		22876	
QY	901	ATCTCTCTAAGCTTGGGGTGTGTACAGACAGGAGGAGATGAGATCTTGGAG		960	
DB	22877	ATCTCTCTAAGCTTGGGGTGTGTACAGACAGGAGGAGATGAGATCTTGGAG		22936	
QY	961	GAAACATGTATCCGGAGAGTTAACGCAATGTGCGGGCATTCAGACTAATATCT		1020	

DB	22937	GAAACATGTATCCGGAGAGTTAACGCAATGTGCGGGCATTCAGACTAATATCT		22996	
QY	1021	GTCGGGAGCAGAGCTTGAAGCAAGATGATTAACAATTTTGGCACCCCCACTG		1080	
DB	22997	GTCGGGAGCAGAGCTTGAAGCAAGATGATTAACAATTTTGGCACCCCCACTG		23056	
QY	1081	CTCGCCCTCTTGAAGAGATGGAATTGGCAGGCAAGAGAGTTCTGTCAGACA		1140	
DB	23057	CTCGCCCTCTTGAAGAGATGGAATTGGCAGGCAAGAGAGTTCTGTCAGACA		23116	
QY	1141	CAGCTGTGAGAGTTTCTGACGTGAGAGATGATGATTAATGCTTCTGAGAGA		1200	
DB	23117	CAGCTGTGAGAGTTTCTGACGTGAGAGATGATGATTAATGCTTCTGAGAGA		23176	
QY	1201	ATGATTTCCGATGCTTTGAGATGATTTCTGATACGACCTTGAAGTATTAAC		1260	
DB	23177	ATGATTTCCGATGCTTTGAGATGATTTCTGATACGACCTTGAAGTATTAAC		23236	
QY	1261	ATGAGAGTCAAAAGAGCATGATCACAGAAAGCAATAGTGTCAAAAAGTTCT		1320	
DB	23237	ATGAGAGTCAAAAGAGCATGATCACAGAAAGCAATAGTGTCAAAAAGTTCT		23296	
QY	1321	GCACTTGTAGTACCTTGTGATCGAGCAGATTAATGAACACAGAGCACTGCA		1380	
DB	23297	GCACTTGTAGTACCTTGTGATCGAGCAGATTAATGAACACAGAGCACTGCA		23356	
QY	1381	CAGCTTGTCAAGACGACCTGTGTCCATGCAATTTGATTAACCTGCAACCT		1440	
DB	23357	CAGCTTGTCAAGACGACCTGTGTCCATGCAATTTGATTAACCTGCAACCT		23416	
QY	1441	ATGCGAGGACAAAAGAGCTAGGCGAGTTAAGCTCCATAGAAATGGCTGAAT		1500	
DB	23417	ATGCGAGGACAAAAGAGCTAGGCGAGTTAAGCTCCATAGAAATGGCTGAAT		23476	
QY	1501	AAAAGATCTACAGATGAGAGGCGACTCTGATCTCTTGTGTGAGATTTAAT		1560	
DB	23477	AAAAGATCTACAGATGAGAGGCGACTCTGATCTCTTGTGTGAGATTTAAT		23536	
QY	1561	GGAAGGTTTGGGTGAGATGAAAAGATTAATGAATTTGCTGCTCCCAATG		1620	
DB	23537	GGAAGGTTTGGGTGAGATGAAAAGATTAATGAATTTGCTGCTCCCAATG		23596	
QY	1621	TCATGAATCTAAGACGATGATGAAGACGATTAAGGATGTAATCTTCTGTT		1680	
DB	23597	TCATGAATCTAAGACGATGATGAAGACGATTAAGGATGTAATCTTCTGTT		23656	
QY	1681	TTCTTAAATTTTCTCTTGCATTTCTACTGATCTTAGATGTTCAATTTGAT		1740	
DB	23657	TTCTTAAATTTTCTCTTGCATTTCTACTGATCTTAGATGTTCAATTTGAT		23716	
QY	1741	ATGCGCAACCAAGAGCTGTGGAATTAATGACAGATGAGCTCTTAGAGCA		1800	
DB	23717	ATGCGCAACCAAGAGCTGTGGAATTAATGACAGATGAGCTCTTAGAGCA		23776	
QY	1801	TTCTATGCTCCAGAGGCCATGCTGGAGATGAGTTCTGATGTTAGAGACAG		1860	
DB	23777	TTCTATGCTCCAGAGGCCATGCTGGAGATGAGTTCTGATGTTAGAGACAG		23836	
QY	1861	GGCTATTTGGAGGCCAGACCTCTCACACCGGAGTTAGCTGAGATGGGTTAG		1920	
DB	23837	GGCTATTTGGAGGCCAGACCTCTCACACCGGAGTTAGCTGAGATGGGTTAG		23896	
QY	1921	GCTTGGGCTCAGAGCGCATATGTTTCTGAGGTGTTGCCAATCTATAGCT		1980	
DB	23897	GCTTGGGCTCAGAGCGCATATGTTTCTGAGGTGTTGCCAATCTATAGCT		23956	
QY	1981	GCAACGAGCAAGATCGACATTAATCAACATCTCAAGATTTCTCCGCCAAGA		2040	
DB	23957	GCAACGAGCAAGATCGACATTAATCAACATCTCAAGATTTCTCCGCCAAGA		24016	
QY	2041	AATTTGATATATGCTTTAGTTTTCATTTGGAATTTAAAGTTTGTGCTCGT		2100	

Dp	24017	AAATTGATPANA	CGCTTTTGGTTTGGTCATCGAATTTAAAGTTTGTGGTCCGGTTAA	24076
Qy	2101	TGCACTGTTATGTA	TATATATCTATGATTTATAGCGAAACAAAGGCTAAATTTGAGTTG	2160
Dp	24077	TGATCTGTTATGTA	TATATCTATGATTTATAGCGAAACAAAGGCTAAATTTGAGTTG	24136
Qy	2161	AAATCATACCA	AGAGATGCTGTAAGGAGCTGAGGCAAGATCTCTGAGGACATACGACG	2220
Dp	24137	AAATCATACCA	AGAGATGCTGTAAGGAGCTGAGGCAAGATCTCTGAGGACATACGACG	24196
Qy	2221	CTGAATCTACTT	TAAAGAACAAAGCTCTCAAAACGAAACAGACATCCAGTGTGAGGAA	2280
Dp	24197	CTGAATCTACTT	TAAAGAACAAAGCTCTCAAAACGAAACAGACATCCAGTGTGAGGAA	24256
Qy	2281	TCCTGGAAATTT	TGAGCCGAAAGCTGCGTAGAATCGACAGAGATATTCGATCGTAGA	2340
Dp	24257	TCCTGGAAATTT	TGAGCCGAAAGCTGCGTAGAATCGACAGAGATATTCGATCGTAGA	24316
Qy	2341	CAGAGAACTA	AGATGCAATGACATGAAACAAACAGGAAAGAGATGATTTTCTAGAAAT	2400
Dp	24317	CAGAGAACTA	AGATGCAATGACATGAAACAAACAGGAAAGAGATGATTTTCTAGAAAT	24376
Qy	2401	CACAAACTTGA	CAATTTTGATTTACCTATGATTCACATTTTGGATTTAATTTGTCACAA	2460
Dp	24377	CACAAACTTGA	CAATTTTGATTTACCTATGATTCACATTTTGGATTTAATTTGTCACAA	24436
Qy	2461	AAAACTGTGTG	TGTAAGATGATGACACAGACAGGTTTTTCAATGATTTCAATCAA	2520
Dp	24437	AAAACTGTGTG	TGTAAGATGATGACACAGACAGGTTTTTCAATGATTTCAATCAA	24496
Qy	2521	CAGATCCATGA	AGAGAGACGCAAAAGGAGAGAAATTTGGAATGTTGACAGAGAGAGAA	2580
Dp	24497	CAGATCCATGA	AGAGAGACGCAAAAGGAGAGAAATTTGGAATGTTGACAGAGAGAGAA	24556
Qy	2581	CGTCCAAAGGT	TGTTGGCCAGCAGCAGCAAGACATTAATCCCTCTAGCATGACGATTC	2640
Dp	24557	CGTCCAAAGGT	TGTTGGCCAGCAGCAGCAAGACATTAATCCCTCTAGCATGACGATTC	24616
Qy	2641	CGAAAGGGTA	TATGTACTAATTAATCCCTCGGCGTTTTTGTTTTCAAACTTA	2700
Dp	24617	CGAAAGGGTA	TATGTACTAATTAATCCCTCGGCGTTTTTGTTTTCAAACTTA	24676
Qy	2701	AGAGTAATCTA	ATTAATTCGCGTTTTTGAATCTTTGCGAGAGCTGAGAGATGTCAGGCTTC	2760
Dp	24677	AGAGTAATCTA	ATTAATTCGCGTTTTTGAATCTTTGCGAGAGCTGAGAGATGTCAGGCTTC	24736
Qy	2761	ATCAGAGTTT	CAGAAGAAAGATGAGAGATTTGTGAAAGAGAGGAGATGCTGATAAA	2820
Dp	24737	ATCAGAGTTT	CAGAAGAAAGATGAGAGATTTGTGAAAGAGAGGAGATGCTGATAAA	24796
Qy	2821	GATCAAGAGAA	GATGGAAGACATGAAAGAAAGGACATCAGAGAGATTTTGAATCTTG	2880
Dp	24797	GATCAAGAGAA	GATGGAAGACATGAAAGAAAGGACATCAGAGAGATTTTGAATCTTG	24856
Qy	2881	GAGAAAGAA	TGTTGATGAGCTTTGGAAACAGCTGATGTAACAGCATGCGCTTCAATGAA	2940
Dp	24857	GAGAAAGAA	TGTTGATGAGCTTTGGAAACAGCTGATGTAACAGCATGCGCTTCAATGAA	24916
Qy	2941	GATGATTTGA	CAAAAAGTCTGATCACAGAACAATTAATGTTCTTTGTGTTTGTCTTTG	3000
Dp	24917	GATGATTTGA	CAAAAAGTCTGATCACAGAACAATTAATGTTCTTTGTGTTTGTCTTTG	24976
Qy	3001	GTAATGTGGA	AAAGTGAAGATCTGAGAGATCCCATTTAAATCTAGAACAAATCTAAGGA	3060
Dp	24977	GTAATGTGGA	AAAGTGAAGATCTGAGAGATCCCATTTAAATCTAGAACAAATCTAAGGA	25036
Qy	3061	GATTATAGAT	TATATATCTCCCAATTTTATGAGACGAGATCTAAGAGCAATTAAGTTCTT	3120
Dp	25037	GATTATAGAT	TATATATCTCCCAATTTTATGAGACGAGATCTAAGAGCAATTAAGTTCTT	25096
Qy	3121	GTCATTAATA	ACCAAGTTTCCCTTATGATTTTGTTTTGTGTAATTTGCATGAAAG	3180
Dp	25097	GTCATTAATA	ACCAAGTTTCCCTTATGATTTTGTTTTGTGTAATTTGCATGAAAG	25156

**RESULT 3**  
**LOCUS** AF239719  
**DEFINITION** Arabidopsis thaliana SGS3 gene, complete cds.  
**ACCESSION** AF239719  
**VERSION** AF239719.1 GI:8164029  
**KEYWORDS** Arabidopsis thaliana (thale cress)  
**ORGANISM** Arabidopsis thaliana  
**REFERENCE** Arabidopsis thaliana  
**AUTHORS** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroside II; Brassicales; Brassicaceae; Arabidopsis.  
**TITLE** 1 (bases 1 to 2254)  
**JOURNAL** Mourrain, P., Becclin, C., Elmayer, T., Feuerbach, F., Godon, C.,  
**MEDLINE** Morel, J.B., Jonette, D., Lacombe, A.M., Nkic, S., Picault, N.,  
**PUBMED** Remoue, K., Sanial, M., Vo, T.A. and Vaucheret, H.  
**REFERENCE** Arabidopsis SGS2 and SGS3 genes are required for  
**AUTHORS** posttranscriptional gene silencing and natural virus resistance  
**TITLE** Cell 101 (5), 533-542 (2000)  
**FEATURES**  
**source** Submitted (22-FEB-2000) Biologie Cellulaire, INRA, Route de  
**source** Saint-Cyr, Versailles 78026, France  
**location/Qualifiers**  
**1**. 2254  
**/organism**="Arabidopsis thaliana"  
**/mol\_type**="genomic DNA"  
**/culivar**="Columbia"  
**/db\_xref**="taxon:3702"  
**/chromosome**="V"  
**/join**("  
**/product**="SGS3"  
**/join**(1..963,1037..1328,1440..1684,1787..1953,2044..2254)  
**/codon\_start**=1  
**/product**="SGS3"  
**/protein\_id**="AAF73960.1"  
**/db\_xref**="GI:8164030"  
**/translation**="MSSRAGPMSSKKEKNVQGYRPEVQLVQGLAGTRLASODDGGEM  
EVIISKNNKRGNTSGKTWVNSQNSPPRPMGQOQGRSGNSGRGNVSGRGNGRG  
IOANISGRALSRKYDNNPVA PVPVSPLEGGNMGVAGSAOHTAVQEPDVED  
VDNASEBENDSDALDDSDPDLASDYPDSDSGSKSRKONWPKFPGSLDLSIEO  
INEPDRHCPACQNGPGLATIDNINLHLAARTGARGRYKLHRELAETVLEDMRG  
ASVIPCGBEIVQWKGIDEDKDYETVWPPVITINTRLDKDDNDKWLGNGBELTYF  
DKYELARARHSYGPQHRGMSVTFPSSATGYLAEARHLRELAEGWLDRIAWGQKSM  
FSGGROLVYGLATKODLDI FNOHSGQTRLFEIYSYGVEMVVEKELRQISENQOQNTY  
FKNTLSKONKRAKVALESEILRETRAEONRLVROETKMOHEONREEMADHDFE  
MDSKIOHRENDKAEENFEMLOQORAVVGOQOONINPSSNDQKRAEVSFTIEF  
QEKEMEFVEEREMILIKQEKMEKMKRHHETIDLEKFPDALEQLMYKIGLNHND  
D"  
**ORIGIN**  
**Query Match** 68.8%; Score 2254; DB 8; Length 2254;  
**Best Local Similarity** 100.0%; Pred. No. 0;  
**Matches 2254; Conservative** 0; Mismatches 0; Indels 0; Gaps 0;  
**QY** 636 ATGAGCTTCAAGGCTGATCCATATCTTAAAGAAAAGACGTTCAAGGCTGTTATAGGCTT 755  
**Db** 1 ATGAGCTTCAAGGCTGATCCATATCTTAAAGAAAAGACGTTCAAGGCTGTTATAGGCTT 60  
**QY** 756 GAGGTTGAACAGTTGCTTCAAGTTTGGACGAGGACGACGCTGCTTCTTCAACAAGTAGAT 815

```
Db 61 GAGGTGAACAGTTGTTCAAGGTTGGCAGGGGACGAGCTGCTTCTTCAACAGATGAT 120
Qy 816 GAGAGAGAGTGGAGGTCATTTTCCAGAGAAACAAGAACCAACAGGAAACCTTTCGA 875
Db 121 GAGAGAGAGTGGAGGTCATTTTCCAGAGAAACAAGAACCAACAGGAAACCTTTCGA 180
Qy 876 AAAACTTGGGTTTCTCAGAAATTCGATCTGAGAGCTTGGGGTGGTGCAGCAGCAGG 935
Db 181 AAACTTGGGTTTCTCAGAAATTCGATCTGAGAGCTTGGGGTGGTGCAGCAGCAGG 240
Qy 936 AGAGTAGCAACGTCATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGGCAATGCT 995
Db 241 AGAGTAGCAACGTCATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGGCAATGCT 300
Qy 996 CCGGGGCTTCAAGCTAATATCTGGTCCGGGAGCAGAGCTTGGAGCAAAAATGATGATAC 1055
Db 301 CCGGGGCTTCAAGCTAATATCTGGTCCGGGAGCAGAGCTTGGAGCAAAAATGATGATAC 360
Qy 1056 AACTTGTGGCACCCCACTGATCTGGCCCTCTTGGAAAGGAGATGGAATTTGGCAG 1115
Db 361 AACTTGTGGCACCCCACTGATCTGGCCCTCTTGGAAAGGAGATGGAATTTGGCAG 420
Qy 1116 GCAAGAGAGGTTCTGCTCAGCAACAAGCTGTGAGAGATTTCTGACGTGAGAGATGAT 1175
Db 421 GCAAGAGAGGTTCTGCTCAGCAACAAGCTGTGAGAGATTTCTGACGTGAGAGATGAT 480
Qy 1176 GTGATATATGCTTCTGAGAGAGAGATGATTCGATCTTGGATGATTTGATGACAC 1235
Db 481 GTGATATATGCTTCTGAGAGAGAGATGATTCGATCTTGGATGATTTGATGACAC 540
Qy 1236 CTGCAAGTATGATTAATGATCTGGAGTGAATGTCAAAGAGCAATGATACGAAACAG 1295
Db 541 CTGCAAGTATGATTAATGATCTGGAGTGAATGTCAAAGAGCAATGATACGAAACAG 600
Qy 1296 AATAGTGTCTCAAAAAGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGAT 1355
Db 601 AATAGTGTCTCAAAAAGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGAT 660
Qy 1356 GAAACACAGAGGAGTGGCATTGTCAGCTTGTGACAAAGAGCTGGTGCATGATGAG 1415
Db 661 GAAACACAGAGGAGTGGCATTGTCAGCTTGTGACAAAGAGCTGGTGCATGATGAG 720
Qy 1416 TATACCTGACCTCTACTAGCTCAGGAGCAAAAAGAGCTAGGAGATTAAGCTC 1475
Db 721 TATACCTGACCTCTACTAGCTCAGGAGCAAAAAGAGCTAGGAGATTAAGCTC 780
Qy 1476 CATAGAGATTTGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGCGCATCTGCTAT 1535
Db 781 CATAGAGATTTGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGCGCATCTGCTAT 840
Qy 1536 CCTTGTGGTGAATTTATGGGAGTGGAGAGGTTTGGGTGAGATGAAGATTAAGAA 1595
Db 841 CCTTGTGGTGAATTTATGGGAGTGGAGAGGTTTGGGTGAGATGAAGATTAAGAA 900
Qy 1596 ATTGTGCGGCTCCAAATGCTCATCATGATGATGATGATGATGATGATGATGATGAT 1655
Db 901 ATTGTGCGGCTCCAAATGCTCATCATGATGATGATGATGATGATGATGATGATGAT 960
Qy 1656 AAGGTGAATTTCTTGTCTTTTACTTTTAAATTTTCTTGAATTTCTACTGATCTTA 1715
Db 961 AAGGTGAATTTCTTGTCTTTTACTTTTAAATTTTCTTGAATTTCTACTGATCTTA 1020
Qy 1716 GAATGTTTCACTTGTAGTGGCTGGCAGTGGCAACAAGAGCTGCTGAATTTCTTGA 1775
Db 1021 GAATGTTTCACTTGTAGTGGCTGGCAGTGGCAACAAGAGCTGCTGAATTTCTTGA 1080
Qy 1776 GTATGAGGCTCTTGAAGACGCAATTCATGATGCAACAGGCGCATGCTGGAGTGAATG 1835
Db 1081 GTATGAGGCTCTTGAAGACGCAATTCATGATGCAACAGGCGCATGCTGGAGTGAATG 1140
Qy 1836 TCTGATGTTTGAAGACGCAATTCATGATGCAACAGGCGCATGCTGGAGTGAATG 1895
Db 1141 TCTGATGTTTGAAGACGCAATTCATGATGCAACAGGCGCATGCTGGAGTGAATG 1200
Qy 1896 AGCTGAGATGAGGTTAGATAGAAATTCGCTGGGTTCAGAACGAGATGATTTTCTGAGG 1955
Db 1201 AGCTGAGATGAGGTTAGATAGAAATTCGCTGGGTTCAGAACGAGATGATTTTCTGAGG 1260
Qy 1956 TGTTCGCAACTGATATGAGCTTCCCTTGGCAAGAGCAAGATCTGGAACATATTCATCA 2015
Db 1261 TGTTCGCAACTGATATGAGCTTCCCTTGGCAAGAGCAAGATCTGGAACATATTCATCA 1320
Qy 2016 CTCTCAAGGTTCTCTCCCAAGAAATTTGATATATGCTTTTATGTTTGTCAATGGAAT 2075
Db 1321 CTCTCAAGGTTCTCTCCCAAGAAATTTGATATATGCTTTTATGTTTGTCAATGGAAT 1380
Qy 2076 TTTAAATTTTGTGTCCTGCTTAAAGCATCTGTATGATATATCTATGATTTATGAGG 2135
Db 1381 TTTAAATTTTGTGTCCTGCTTAAAGCATCTGTATGATATATCTATGATTTATGAGG 1440
Qy 2136 CAAAACAAAGCTGAATTTGAGTTGAATCATACCAAGAGATGATGTTAAAGAGCTGAG 2195
Db 1441 CAAAACAAAGCTGAATTTGAGTTGAATCATACCAAGAGATGATGTTAAAGAGCTGAG 1500
Qy 2196 GAGATCTCTGAGACCAATCAGCAGCTGAACTATTTAAGAACAGCTCTCAAAACAGAA 2255
Db 1501 GAGATCTCTGAGACCAATCAGCAGCTGAACTATTTAAGAACAGCTCTCAAAACAGAA 1560
Qy 2256 CAAACACGCAAGCTGTTAGAGATCTCTGAGAAATTTATGAGGAGAAAGCTGGTGAAC 2315
Db 1561 CAAACACGCAAGCTGTTAGAGATCTCTGAGAAATTTATGAGGAGAAAGCTGGTGAAC 1620
Qy 2316 TGCAGAGATATTCGATCTGAGACAGAGAACTAAGATCAGACATGAAACAGAA 2375
Db 1621 TGCAGAGATATTCGATCTGAGACAGAGAACTAAGATCAGACATGAAACAGAA 1680
Qy 2376 AGAGTATGATTTTCTTGAAGAAATCAAACTTGAATTTTATTAATCTACTGATGTA 2435
Db 1681 AGAGTATGATTTTCTTGAAGAAATCAAACTTGAATTTTATTAATCTACTGATGTA 1740
Qy 2436 CATTTTGATTAATTTGTCCAAACAAAACCTGTGGTGGTTGAAGATGATGATGATGAT 2495
Db 1741 CATTTTGATTAATTTGTCCAAACAAAACCTGTGGTGGTTGAAGATGATGATGATGAT 1800
Qy 2496 CAGTTTTTCAATGATTCATCAATCAAGATCCATGAAAGAGAGACGCAAGAGAGAA 2555
Db 1801 CAGTTTTTCAATGATTCATCAATCAAGATCCATGAAAGAGAGACGCAAGAGAGAA 1860
Qy 2556 TTTGAGATGTTGCAAGCAGAGAAAGTGCAGAGTTGTTGGCCAGCAGCAGCAGAAAT 2615
Db 1861 TTTGAGATGTTGCAAGCAGAGAAAGTGCAGAGTTGTTGGCCAGCAGCAGCAGAAAT 1920
Qy 2616 TAATCCCTCTAGCAATGAGATTTGCCGAAAGAGGTATATGATTAATTAATTAATCCCT 2675
Db 1921 TAATCCCTCTAGCAATGAGATTTGCCGAAAGAGGTATATGATTAATTAATTAATCCCT 1980
Qy 2676 CTGGCGTTTTTGTCTTCAAACTTAAGATGATGATGATGATGATGATGATGATGATGAT 2735
Db 1981 CTGGCGTTTTTGTCTTCAAACTTAAGATGATGATGATGATGATGATGATGATGATGAT 2040
Qy 2736 CAGAGCTGAGAGATGCTCAAGCTTCAATCAGATTTCAAGAGAAAGAGATGAGAGATTTGT 2795
Db 2041 CAGAGCTGAGAGATGCTCAAGCTTCAATCAGATTTCAAGAGAAAGAGATGAGAGATTTGT 2100
Qy 2796 GGAAGAGAGGAGATGCTGATTAAGATCAAGAGAAAGATGAGAGATGAGAGAGAG 2855
Db 2101 GGAAGAGAGGAGATGCTGATTAAGATCAAGAGAAAGATGAGAGATGAGAGAGAG 2160
Qy 2856 GCATCAAGAGAGATTTTATGATCTGAGAGAAAGATTTTATGATGAGGCTTTGGAACAGCTCAT 2915
Db 2161 GCATCAAGAGAGATTTTATGATCTGAGAGAAAGATTTTATGATGAGGCTTTGGAACAGCTCAT 2220
Qy 2916 GTTCAAGCATGGCTTCAACATGAGATGATGA 2949
Db 2221 GTTCAAGCATGGCTTCAACATGAGATGATGA 2254
```



RESULT 4  
BT002944  
LOCUS  
DEFINITION  
Arabidopsis thaliana clone RAF14-93-K05 (R20243) unknown protein  
(At5g23570) mRNA, complete cds.  
BT002944  
ACCESSION  
BT002944.1 GI:27754622  
KEYWORDS  
FLI CDNA.  
ORGANISM  
Arabidopsis thaliana (thale cress)  
Bukeriyota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 2162)  
REFERENCE  
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,  
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamliya, A., Kawai, J., Kim, C.J., Narusaka, M.,  
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,  
Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Arabidopsis full length cDNA clones  
Unpublished  
2 (bases 1 to 2162)  
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,  
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamliya, A., Kawai, J., Kim, C.J., Narusaka, M.,  
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,  
Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Direct Submission  
Submitted (15-JUN-2003) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL CDNA: 'RIKEN  
Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamliya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEN (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Yamada, K., Chan, M.M.,  
Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S.,  
Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G.,  
Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M.,  
Palm, C.J., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W.,  
Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEN) and Seki, M. (RIKEN GSC) contributed equally to  
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEN)  
contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome  
submitted to Genbank.

FEATURES  
source  
1..2162  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAF14-93-K05 (R20243)"  
/ecotype="Columbia"  
/note="This clone is in a modified pBluescript vector2  
(lambda ps) as a BamHI/XhoI insert."  
1..2162  
/gene="At5g23570"  
1..91  
/gene="At5g23570"  
92..1969  
/gene="At5g23570"  
/codon\_start=1

ORIGIN  
Query Match 29.7%; Score 972; DB 8; Length 2162;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
misc\_difference 2146  
/evidence=experimental  
/product="unknown protein"  
/protein\_id="AA02757.1"  
/db\_xref="GI:27754623"  
/translation="MSSRAGPMKSKKNGVGRPEVQIYVGLAGTRLASDQDGEW  
EVI SKNKKRGTSTKTVSVSONSNPPRMGGQQQQRGSNVSGRGNVSGRGNRG  
IQANISGRGRLSKTDNNFVAPVPVSRPPLGGMMWQARGSAGHTVAQVDFPVEVD  
VDNASEEDSDALDDSDDLASDDVDVDSQSKSGSRKQNMFKKFFGSLSLSEQ  
INBPORWMCPCACONGPGAIDWNLNHLPLAHRATKARERKLHRELAEVLEKLOMRG  
ASVIPCGLYGGWKGDEKDYELVWPPVLIIMTRLDKNDNDKWLGMGNELLEYF  
DKREALRARSYVGPOGRGASVLMFESSATGLVLEARHLRELAEVLEKLOMRG  
FSGGVRLYGLFLATRODLIPNHSQGRTRLEFKLSVQMTKQELROISENQOLNT  
FKNKL SKNKKHAKVLEESLFI MSEKTRAEEDNRIVROTKQHEQNRKEMADHDF  
WDSIKQIHRDRDKENFEMLDQQRKAVVQOQOINPSSDDCKRKAEEVSSFLTF  
QEKMEEFVEEREMLIKQDEKMEKMKHHEIIPLEKEPDEALQLMYKHLNHED  
D"  
1970..2162  
/gene="At5g23570"  
/note="not present in genomic sequence"

QY 687 GTTGAATAATGAGTTCTAGGGCTGTGCTCAATGTCTAAGAAAGAACCTTCAGGGTGT 746  
DB 83 GTTGAATAATGAGTTCTAGGGCTGTGCTCAATGTCTAAGAAAGAACCTTCAGGGTGT 142  
QY 747 TATAGGCTGAGGTTGAAGAGTTGTTCAAGTTTGGCAGGACGAGATGCTTCTTCA 806  
DB 143 TATAGGCTGAGGTTGAAGAGTTGTTCAAGTTTGGCAGGACGAGATGCTTCTTCA 202  
QY 807 CAAGATGATGAGGAGAGAGTGGAGGTCATTTCAAGAAAGAACCAAGAAAC 866  
DB 203 CAAGATGATGAGGAGAGAGTGGAGGTCATTTCAAGAAAGAACCAAGAAAC 262  
QY 867 ACTTCTGAAAAAATCTGGGTTTCTCAAGATTCTCTCTTGAAGCTTGGGTTGCTG 926  
DB 263 ACTTCTGAAAAAATCTGGGTTTCTCAAGATTCTCTCTTGAAGCTTGGGTTGCTG 322  
QY 927 CAGCAAGGAGAGGATGACCAAGTATCTGGAGAGAAACAATGTTCCGGAGAGAGTAA 986  
DB 323 CAGCAAGGAGAGGATGACCAAGTATCTGGAGAGAAACAATGTTCCGGAGAGAGTAA 382  
QY 987 GGCATGATGTCGGGGCATTCACATCAATCTGTCGGGGGACGAGCGTTGAGCAGAAAG 1046  
DB 383 GGCATGATGTCGGGGCATTCACATCAATCTGTCGGGGGACGAGCGTTGAGCAGAAAG 442  
QY 1047 TATGATTAACAATTTGTGGACACCCCACTGTATCTGCGCTCTTGGAGAGAGATGG 1106  
DB 443 TATGATTAACAATTTGTGGACACCCCACTGTATCTGCGCTCTTGGAGAGAGATGG 502  
QY 1107 AATTGGCAGGCAAGAGAGTCTGCTCAGCAACACAGCTGTGACAGAGTTTCTGACGTG 1166  
DB 503 AATTGGCAGGCAAGAGAGTCTGCTCAGCAACACAGCTGTGACAGAGTTTCTGACGTG 562  
QY 1167 GAGGATGATGTGATTAATGCTTCTGAGAGAGAAATGATTCGATCTTGGATGATTTCT 1226  
DB 563 GAGGATGATGTGATTAATGCTTCTGAGAGAGAAATGATTCGATCTTGGATGATTTCT 622  
QY 1227 GATGACGACCTTGGCAAGATGATTAATGATCTCGATGATGATCAAAAGACCATGATCA 1286  
DB 623 GATGACGACCTTGGCAAGATGATTAATGATCTCGATGATGATCAAAAGACCATGATCA 682  
QY 1287 CGAAGACGAATTAATGATGATCAAAAGCTTCTTGGCAGCTTGGATGATGATGATGAG 1346  
DB 683 CGAAGACGAATTAATGATGATCAAAAGCTTCTTGGCAGCTTGGATGATGATGATGAG 742  
QY 1347 CAGATTAATGAACACAGAGCGATGTCATGCTTGTCAAGACGACCTGTGCTCC 1406

Db	743	CAGATAAATGAAACACAGAGGAGTGGCAATTTGTCAGGTTTGCAAAAGGACCTTGTCGC	802
QY	1407	ATCGATTGGTAAACCTGCAACCTCTACTACTGTCATGCGAGACAAAGAGCTTAGCGCA	1466
Db	803	ATCGATTGGTAAACCTGCAACCTCTCTACTACTGTCATGCGAGACAAAGAGAGCTTAGCGCA	862
QY	1467	GTTAAGCTCCATAGAGATTGGCTGAGATTTTAGAAAAGATCTACAGATGAGAGCGCA	1522
Db	863	GTTAAGCTCCATAGAGATTGGCTGAGATTTTAGAAAAGATCTACAGATGAGAGCGCA	922
QY	1527	TCTGTCATTCCTGTGGTGAATTAATGGCAGTCGGAAGGTTGGGTGAGAGTCAAAAG	1586
Db	923	TCTGTCATTCCTGTGGTGAATTAATGGCAGTCGGAAGGTTGGGTGAGAGTCAAAAG	982
QY	1587	GATTATGAAATTGCTGGCCTCCAAATGTCATCATCAGAAATACTAGACTGGATMAAGAC	1646
Db	983	GATTATGAAATTGCTGGCCTCCAAATGTCATCATCAGAAATACTAGACTGGATMAAGAC	1042
QY	1647	GATAACGATTAAG	1658
Db	1043	GATAACGATTAAG	1054

RESULT 5			
AX078761			
LOCUS	AX078761	1878 bp	DNA
DEFINITION	Sequence 2 from Patent WO010551.		linear
ACCESSION	AX078761		
VERSION	AX078761.1	GI:13158380	
KEYWORDS			
SOURCE			
ORGANISM	Arabidopsis thaliana ('hafe cress)		
	Arabidopsis thaliana		

1  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Beclin, C., Elmayer, T. and Vaucheret, H.  
Novel sg33 plant gene and use thereof  
Patent: WO 0105951-A 2 25-JAN-2001;  
AVENTIS CROSCIENCE S.A. (FR); INSTITUT NATIONAL DE LA RECHERCHE  
AGRONOMIQUE (FR)

FEATURES	
source	location/qualifiers
	1..1878
	/organism="Arabidopsis thaliana"
	/mol_type="unassigned DNA"
	/db_xref="taxon:3702"
CDS	1..1878

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CAC2419.1"
/db_xref="GI:13158381"
/db_xref="REMBRBL:CAC2419"
/translation="MSRRAGPMSEKKEVYGGYREVEQLVGLATGRILASSODDGGEM
EVISSKKNKRGNTSGKTWVSONSPNPRAGGQGGGNSVSGKGNVSRGNGNG
IGANISGRGDAISRYNNVYAPVPVSRPLEGGMMNARGGSAQHTAVQEPVEYD
VDNASEERDSDALDDDDLDADYDSDVSQSHSGRKQKPKFKFGSDLSIDQ
INBPQRQWHPACONGGAGDAIDWYLLHLLHARTKARRYKALHHELAVEIKLQMGQ
ASYAPCEGIYGMQKGLDEDEKYEIVWPVWIIINTRLDXDDNKWGMNOELLEFF
DKYEALEFARSHYGOQGRGMSVLMFESSAQTGAEALHHELMAGMDRLAMQGRSM
FSGGVRQLVGFIAITKODLDIYNHOSKQKTLKFKLSYQMVYKELPQISEDNQILNY
FKKLSKQNHAKYLVESLEIMSEKLRATYADNRIVQRTMGHQRREKMDADRFF
WDSIKQIHERRDAAKEENFEMILQGOERKARVVGQOQNIINPSSNDRCRRAEVSSFTFF
QGEEMEFVEVEREMILIDQEKQKEDMKGRHHEIFLDEKEFDEALBOLMTYKHGLNED
D"

```

```
ORIGIN
Query Match          29.4%; Score 963; DB 6; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches   963; Conservative    0; Mismatches      0; Indels     0; Gaps       0

QY      696 ATGAGTTTCAGGGCTGTCTCAATGTCCTAAGAAGAAACGTTCAAGGTGATTAGSCTT 755
        |||
1 ATGAGTTTCAGGGCTGTCTCAATGTCCTAAGAAGAAACGTTCAAGGTGATTAGSCTT 60
```

QY	756	GAGGTGGAACAGTTGGTCTCAAGGTTTGGCAGAGGACGAGACTGGCTTCTTCAACAAGATGAT	815
Db	61	GAGGTGGAACAGTTGGTCTCAAGGTTTGGCAGAGGACGAGACTGGCTTCTTCAACAAGATGAT	120
QY	816	GGAGGAGAGTGGAGGTCATTTCCAGAGAGAACAAACCAGAGAAACCTTCTGGA	875
Db	121	GGAGGAGAGTGGAGGTCATTTCCAGAGAGAACAAACCAGAGAAACCTTCTGGA	180
QY	876	AAAACTTGGGTTTCTCAGAAATTCGAATTCCTCTAGACTTGGGGTGGTCAGCAGCAAGG	935
Db	181	AAAACTTGGGTTTCTCAGAAATTCGAATTCCTCTAGACTTGGGGTGGTCAGCAGCAAGG	240
QY	936	AGAGGTAGCAACGATCTCGGAGAGGAAACAAATATACCGGGAGAGGTAAACGCAATGCT	995
Db	241	AGAGGTAGCAACGATCTCGGAGAGGAAACAAATATATCCGGGAGAGGTAAACGCAATGCT	300
QY	996	CGGGGCATTCGAAGCTAACATATCTCGGAGGAGAGCGTGGAGCAAAAGTATGATATAC	1055
Db	301	CGGGGCATTCGAAGCTAACATATCTCGGAGGAGAGCGTGGAGCAAAAGTATGATATAC	360
QY	1056	AACCTTGTGCAACCCCAACCTGTATCTCGCCCTCTTGGAGAGGAGATGGAATTTGGCAG	1115
Db	361	AACCTTGTGCAACCCCAACCTGTATCTCGCCCTCTTGGAGAGGAGATGGAATTTGGCAG	420
QY	1116	GCAAGAGAGAGTTCTGCTCAGACACACAGCTGTGAGAGATTTCTTGAACGTGAGAGATGAT	1175
Db	421	GCAAGAGAGAGTTCTGCTCAGACACACAGCTGTGAGAGATTTCTTGAACGTGAGAGATGAT	480
QY	1176	CTGCAATTAATGCTTTGACCAACCAATATGCAATGCGATCTGCTTTCATATTTCTGATGACGAT	1235

[illegible]

RESULT 6	BT004380	1909 bp	mRNA	linear	PLN 14-FEB-2003
LOCUS	BT004380				
DEFINITION	Arabidopsis thaliana clone U20243	unknown protein (At5g23570)	mRNA,		
	complete cds.				
ACCESSION	BT004380				
VERSION	BT004380.1	GI:28393932			

## KEYWORDS

## SOURCE

FLI CDNA.  
Arabidopsis thaliana (chale cress)

## ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.

## AUTHORS

1 (bases 1 to 1909)  
Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.U., Narusaka,M.,  
Nguyen,M., Palm,C.U., Sakurai,T., Satou,M., Seki,M., Shimu,P.,  
Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W.,  
Ecker,J.R. and Theologis,A.  
Arabidopsis Open Reading Frame (ORF) Clones  
Unpublished

## REFERENCE

2 (bases 1 to 1909)  
Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.U., Narusaka,M.,  
Nguyen,M., Palm,C.U., Sakurai,T., Satou,M., Seki,M., Shimu,P.,  
Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W.,  
Ecker,J.R. and Theologis,A.

## TITLE

Direct Submission  
Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA

## COMMENT

The RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL CDNA: 'RIKEN  
Arabidopsis Full-Length CDNA'): Seki,M., Narusaka,M., Ishida,J.,  
Hayashizaki,Y. and Shinzaki,K.

The Salik, Stanford, PGEC (SSP) Consortium members constructed and  
sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,  
Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.U., Nguyen,M.,  
Palm,C.U., Shimu,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,  
Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally  
to this work. Shinzaki,K. (RIKEN GSC) and Theologis,A. (SSP  
/PGEC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome  
submitted to Genbank.

## FEATURES

Location/Qualifiers

1..1909  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="U20243"  
/ecotype="Columbia"  
/note="This clone is in pUNI 51."  
1..1909  
/gene="At5g23570"  
1..1878  
/gene="At5g23570"  
/codon\_start=1  
/evidence="experimental  
/product="unknown protein"  
/protein\_id="AA042374.1"  
/db\_xref="GI:28393933"  
/translation="MSSRAGPMSEKKNVGGYREVQVLVQGLAGTLLASODDGGEM  
EVISSKNKRNKRGNTSGKTWSONSPRANGQOQGGVSGGNNVSGRANGNGG  
IOANIISGRGRKSRKYNNFVAPEPVRPLEGGMMNQAGSGAHOHVAOEPFVED  
VDNASEEDSDALDDSDDDLASDDVDSDVSOXSHGRKONKPKKRFSSLDLSIEO  
INERQWHPACQNGBCAIDWYTHLHLHARKRKARKHRELAEVLEKIDKRG  
ASVIFGEITGOMKGLSEDEKDYIVPFWITLWTLTDLDNDKDKMUGNOELLETF  
DKYELARHSYGGQGRGMSVLMFESSAGYLAELHLEHLEMGIDRLAMQKRSK  
FSGGVRLVYGLATKDDIDIFNHSQKTRTLKELKSYQSMVVELELRISDNQOLNY

## ORIGIN

3' UTR  
FRKLSKONKHAULVEESLEIMSEKLRPTAENRIVRORTKMOHREBMDANDRPF  
MSIKOIHERRDAKNEPFLMOQOQBAKVVGGQOQINPSSNDCKRABEYSSFLBF  
QEKEMEEFVEEREMLIKDEKMEKMDKKHHEIFLEKEFDLEALQMLTKHGLHNE  
D"  
1879..1909  
/gene="At5g23570"

Query Match 29.4%; Score 963; DB 8; Length 1909;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

696 ATGAGTTCTAGGGCTGGTCCATGTCTAAGAGAAAGAGCTTAGGGTGTATAGGCTT 755  
1 ATGAGTTCTAGGGCTGGTCCATGTCTAAGAGAAAGAGCTTAGGGTGTATAGGCTT 60  
756 GAGGTGAACAGTTGGTTCAGGTTTGGCAGGACGAGCTGCTTTCACAAGATGAT 815  
61 GAGGTGAACAGTTGGTTCAGGTTTGGCAGGACGAGCTGCTTTCACAAGATGAT 120  
816 GAGGAGAGTGGGAGGTCTATTCAGAGAACAGAACAGAACAGAACAGAACAGTTCGGA 875  
121 GGAGGAGAGTGGGAGGTCTATTCAGAGAACAGAACAGAACAGAACAGAACAGTTCGGA 180  
876 AAAACTGGGTTTCTGAAATTCCTCCTAGAGCTTGGGGTGTGACGACGAGG 935  
181 AAACTGGGTTTCTGAAATTCCTCCTAGAGCTTGGGGTGTGACGACGAGG 240  
936 AGAGGTAGCAACGTATCTGGAGAGAGAAACAATGTATCCGGAGAGATTAACGCAATGTG 995  
241 AGAGGTAGCAACGTATCTGGAGAGAGAAACAATGTATCCGGAGAGATTAACGCAATGTG 300  
996 CGGGGCAATTAACATATCTGCTCGGGACGAGCGTTGACGCAAAAGTATGATTAAC 1055  
301 CGGGGCAATTAACATATCTGCTCGGGACGAGCGTTGACGCAAAAGTATGATTAAC 360  
1056 AACTTGTGGCACCACCTGATCTCGCCCTCTTGGAGAGAGATGATTAAGGACG 1115  
361 AACTTGTGGCACCACCTGATCTCGCCCTCTTGGAGAGAGATGATTAAGGACG 420  
1116 GCAAGAGAGGTTCTGCTGACGACAGCTGTGACAGAGTTTCTGACGTGAGATGAT 1175  
421 GCAAGAGAGGTTCTGCTGACGACAGCTGTGACAGAGTTTCTGACGTGAGATGAT 480  
1176 GTGATTAATGCTTCTGAGAGAGAGATGATTCGAGTCTTGGATGATTTCTGATGACGAC 1235  
481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGAGTCTTGGATGATTTCTGATGACGAC 540  
1236 CTGCAAGTATGATTAATGATCTCGATGTGAGTCAAAAGAGCATGATCAAGAAAGCAG 1295  
541 CTGCAAGTATGATTAATGATCTCGATGTGAGTCAAAAGAGCATGATCAAGAAAGCAG 600  
1296 AATTAAGTGTCAAAAAGTTCTTGGCAGCTTGGATGATCTTGGATGATGATGATTAAT 1355  
601 AATTAAGTGTCAAAAAGTTCTTGGCAGCTTGGATGATCTTGGATGATGATGATTAAT 660  
1356 GAACCAAGAGAGAGAGTGTCCAGCTTGTCAAAAGAGCATGATGATGATGATTAAT 1415  
661 GAACCAAGAGAGAGTGTCCAGCTTGTCAAAAGAGCATGATGATGATGATTAAT 720  
1416 TATAACTGCAACCTCTAAGTCTGATGAGAGCAAAAGAGCATGATGATGATTAATGATC 1475  
721 TATAACTGCAACCTCTAAGTCTGATGAGAGCAAAAGAGCATGATGATGATTAATGATC 780  
1476 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGTCAAT 1535  
781 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGTCAAT 840  
1536 CCTTGGTGTGAGATTTATGAGGAGTGAAGGTTTGGGTGAGATGAAGAAAGATTTATGA 1595  
841 CCTTGGTGTGAGATTTATGAGGAGTGAAGGTTTGGGTGAGATGAAGAAAGATTTATGA 900  
1596 ATTGTCTGACCTTCATGATGATCATGATTAATGATGATTAAGAGCATTAAGAT 1655



Db 901 ATTTCGGCTCCAAATGTCATCATGATACTAGACTGATTAAGACGATTAACGAT 960  
Qy 1656 AAG 1658  
Db 961 AAG 963

RESULT 7  
ATHS28171

LOCUS 650 bp DNA linear PLN 29-MAR-2003

DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone 157C08.

ACCESSION AJ528171

VERSION AJ528171.1 GI:26796431

KEYWORDS left border; T-DNA flanking sequence.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Crnaud, C., Dehose, R., Pelletier, G., Lepoint, L., Caboche, M., and Lecharny, A.  
T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites  
EMBO Rep. 3 (12), 1152-1157 (2002)

REFERENCE 2 Balzerque, S.  
2 (bases 1 to 650)

JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

REFERENCE 3 Balzerque, S.  
Direct Submission

AUTHORS Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

TITLE PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.inbio.gen.fr>).

FEATURES  
source  
Location/Qualifiers  
1..650  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Mas811lewskija"  
/db\_xref="taxon:3702"  
/clone="157C08"  
/clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
1..650  
/note="T-DNA flanking sequence  
left border"

ORIGIN  
Query Match 16.5%; Score 541; DB 8; Length 650;  
Best Local Similarity 99.8%; Pred. No. 1.9e-296;  
Matches 591; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1428 CCTCTACTAGCTCATGAGACAAAGAGCTTAGGCGAGTTAGCTCCATAGGAATTG 1487  
Db 1 CCTCTACTAGCTCATGAGACAAAGAGCTTAGGCGAGTTAGCTCCATAGGAATTG 60

Qy 1488 GCTGAAGTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGTCATTCTCTGGAGAG 1547  
Db 61 GCTGAAGTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGTCATTCTCTGGAGAG 120

Qy 1548 ATTATGGGACAGTGAAGGGGTTGGGTGAGAGTGAAGATTATGAATTGTCTGGCCT 1607

Db 121 ATTATGGGACAGTGAAGGGGTTGGGTGAGAGTGAAGATTATGAATTGTCTGGCCT 180  
Qy 1608 CCAATGGCTCAATCATGAAATACATAGACTGGAATAGGACGATTAAGTGAATTTC 1667  
Db 181 CCAATGGCTCAATCATGAAATACATAGACTGGAATAGGACGATTAAGTGAATTTC 240

Qy 1668 TTCTGTCTTTTACTCTCTTTATTTTCTCTGTGATCTAGATCTTGAATTTACAT 1727  
Db 241 TTCTGTCTTTTACTCTCTTTATTTTCTCTGTGATCTAGATCTTGAATTTACAT 300

Qy 1728 GTAGTGGCTCGGACATGGGACCAACGAGCTGCTGGAATCTTGACAAATGAGGCTCT 1787  
Db 301 GTAGTGGCTCGGACATGGGACCAACGAGCTGCTGGAATCTTGACAAATGAGGCTCT 360

Qy 1788 TAGAGCAGCGCATTTCCATAGTGTCCACGAGGCGCATTCGGGATAGTGTCTGATTTGA 1847  
Db 361 TAGAGCAGCGCATTTCCATAGTGTCCACGAGGCGCATTCGGGATAGTGTCTGATTTGA 420

Qy 1848 GAGCAGTGCACATGCTGATTTTGGAGGCGGACGCTCCACCGGAGTTAGCTGAGATGG 1907  
Db 421 GAGCAGTGCACATGCTGATTTTGGAGGCGGACGCTCCACCGGAGTTAGCTGAGATGG 480

Qy 1908 GTTAGATGAATTTGCTGCGGCTCAGAAAGCCGATATTTTCTGAGGCTTGCCCACT 1967  
Db 481 GTTAGATGAATTTGCTGCGGCTCAGAAAGCCGATATTTTCTGAGGCTTGCCCACT 540

Qy 1968 GTATGGCTCTCTTGCAACGACGAGATCTGACATATTCATCAACACTCT 2019  
Db 541 GTATGGCTCTCTTGCAACGACGAGATCTGACATATTCATCAACACTCT 592

RESULT 8  
ATHS27741

LOCUS 45 bp DNA linear PLN 29-MAR-2003

DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone 146F03.

ACCESSION AJ527741

VERSION AJ527741.1 GI:26796001

KEYWORDS left border; T-DNA flanking sequence.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Crnaud, C., Dehose, R., Pelletier, G., Lepoint, L., Caboche, M., and Lecharny, A.  
T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites  
EMBO Rep. 3 (12), 1152-1157 (2002)

REFERENCE 2 Balzerque, S.  
2 (bases 1 to 45)

JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

REFERENCE 3 Balzerque, S.  
Direct Submission

AUTHORS Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

TITLE PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.inbio.gen.fr>).

FEATURES  
source  
Location/Qualifiers  
1..45  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"

```

misc_feature
/clone_1lib="Arabidopsis thaliana T-DNA insertion lines"
1. .45
/clone_1lib="Arabidopsis thaliana T-DNA insertion lines"
left border"

ORIGIN
Query Match 1.4%; Score 45; DB 8; Length 45;
Best Local Similarity 100.0%; Pred. No. 5.2e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2033 CCGAAGAAATTGATATATGCTTATGTTGCTATTCGATTT 2077
1 CCGAAGAAATTGATATATGCTTATGTTGCTATTCGATTT 45

RESULT 9
AC102284/c
LOCUS
DEFINITION
AC102284 170765 bp DNA linear HTG 31-OCT-2002
pieces.
AC102284 GI:24431710
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
MUS musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 170765)
Mus musculus, clone RP24-329L16
Unpublished
2 (bases 1 to 170765)
Birren,B., Linton,N., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Barna,N., Baetien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chopel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heatord,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamaze,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meltrin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,U., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS

```

TITLE  
 JOURNAL  
 COMMENT  
 Submitted (31-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 31, 2002 this sequence version replaced gi:22381142.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MITR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L18299  
 Center clone name: 329 L 16  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 169312 bases at least Q40  
 Consensus quality: 170102 bases at least Q30  
 Consensus quality: 170295 bases at least Q20  
 Insert size: 16300; agarose-fp  
 Insert size: 170365; sum-of-contigs  
 Quality coverage: 17.6 in Q20 bases; agarose-fp  
 Quality coverage: 16.8 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 76853: contig of 76853 bp in length
* 76854 76953: gap of 100 bp
* 78954 78949: contig of 1996 bp in length
* 78950 79049: gap of 100 bp
* 79050 92772: contig of 13723 bp in length
* 92773 92872: gap of 100 bp
* 92873 150873: contig of 58001 bp in length
* 150874 150973: gap of 100 bp
* 150974 170765: contig of 19792 bp in length.
Location/Qualifiers
1. 170765
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-329L16"
/clone_lib="RPCT-24 Male Mouse BAC"
1. 76853
/feature="assembly_fragment"
clone_end:SP6
vector_side:left"
76954 78949
/feature="assembly_fragment"
79050 92772
/feature="assembly_fragment"
92873 150873
/feature="assembly_fragment"
150974 170765
/feature="assembly_fragment"
vector_side:right"

```

ORIGIN

Query Match 0.7%; Score 24; DB 2; Length 170765;  
Best Local Similarity 100.0%; Pred. No. 0.51;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACAACAAACAAATTTAGCAAG 25  
|||||  
DB 65301 ACAACAAACAAATTTAGCAAG 65278

RESULT 10  
AF057056/c 554 bp DNA linear INV 19-DEC-2003  
LOCUS Uroleccon erigeronensis NADH dehydrogenase subunit 1-like protein  
DEFINITION (ND1) gene, partial sequence; mitochondrial gene for mitochondrial product.  
ACCESSION AF057056  
VERSION AF057056.1 GI:8745205  
KEYWORDS mitochondrial Uroleccon erigeronensis  
SOURCE Uroleccon erigeronensis  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrostiphini; Uroleccon.  
REFERENCE 1 (bases 1 to 554)  
Morgan, N.A., Kaplan, M.E., Gelsey, M.J., Murphy, T.G. and Scholes, E.A. Phylogenetics and evolution of the aphid genus Uroleccon based on mitochondrial and nuclear DNA sequences  
JOURNAL Syst. Entomol. 24 (1), 85-93 (1999)  
REFERENCE 2 (bases 1 to 554)  
Morgan, N.A., Kaplan, M.E., Gelsey, M.J., Murphy, T.G. and Scholes, E. Iii.  
TITLE Direct Submission  
JOURNAL Submitted (03-APR-1998) Ecology and Evolutionary Biology, University of Arizona, Biological Sciences West, Tucson, Arizona 85721, USA

FEATURES  
source 1..554  
/organism="Uroleccon erigeronensis"  
/organella="mitochondrion"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:87314"  
<1..>554  
/gene="ND1"  
/gene="ND1"  
misc\_feature 1..554  
/gene="ND1"  
/note="similar to NADH dehydrogenase subunit 1"

ORIGIN  
Query Match 0.7%; Score 23; DB 3; Length 554;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3151 GTTTTGTGTAATAATTCAT 3173  
|||||  
DB 468 GTTTTGTGTAATAATTCAT 446

RESULT 11  
BX004860 101080 bp DNA linear HTG 14-DEC-2002  
LOCUS Danio rerio clone BUSM1-144B24, WORKING DRAFT SEQUENCE, 4 unordered pieces.  
ACCESSION BX004860  
VERSION BX004860.1 GI:26985650  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
REFERENCE 1 (bases 1 to 101080)  
Burton, J.  
TITLE Direct Submission

JOURNAL Submitted (10-DEC-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk  
COMMENT Clone requests: clonerequests@sanger.ac.uk  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zface@sanger.ac.uk  
----- Project Information  
Center project name: dz144B24  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 99769 bases at least Q40  
Consensus quality: 100510 bases at least Q30  
Insert size: 100780; sum-of-coverage  
Insert size: 107197; 9.7% error; agarose-fp  
Quality coverage: 5.75x in Q20 bases; sum-of-coverage Quality coverage: 5.56x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 34070: contig of 34070 bp in length  
\* 34071 34170: gap of 100 bp  
\* 34171 67630: contig of 33460 bp in length  
\* 67631 67731: gap of 100 bp  
\* 67731 81714: contig of 13984 bp in length  
\* 81715 81814: gap of 100 bp  
\* 81815 101080: contig of 19266 bp in length.  
Location/Qualifiers  
1..101080  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="BUSM1-144B24"  
/clone\_1id="BUSM1"  
1..34070  
/note="assembly fragment: 00417  
fragment chain: 1  
clone\_end: T7  
vector side: left"  
34171..67630  
/note="assembly fragment: 00598  
fragment chain: 1"  
67731..81714  
/note="assembly fragment: 00262  
fragment chain: 1"  
81815..101080  
/note="assembly fragment: 00119  
fragment chain: 1  
clone\_end: SP6  
vector side: right"

ORIGIN  
Query Match 0.7%; Score 23; DB 2; Length 101080;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 AATGAACAAATCAACTGA 118  
|||||  
DB 39791 AATGAACAAATCAACTGA 39769

RESULT 12  
AL845510/c 101241 bp DNA linear VRT 30-JAN-2003  
LOCUS AL845510

**DEFINITION** Zebrafish DNA sequence from clone BUSM1-132M23 in linkage group 7, complete sequence.

**ACCESSION** AL845510

**VERSION** AL845510.8 GI:28172206

**KEYWORDS** HTG.

**SOURCE** Danio rerio (zebrafish)

**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

**REFERENCE** Bates, K. 1 (bases 1 to 101241)

**AUTHORS** Direct Submission

**JOURNAL** Submitted (29-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk; On Jan 30, 2003 this sequence version replaced gi:27801659.

**COMMENT** ----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep Repeat names beginning 'Dr' were identified by the Repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.Projects/D\_rerio/fishmask.shtml BUSM1-132M23 is from a zebrafish PAC library  
VECTOR: pCYPAC-6

**FEATURES** Location/Qualifiers

**Source** 1..101241  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="BUSM1-132M23"  
/clone\_lib="BUSM1"

**ORIGIN**

Query Match 0.7%; Score 23; DB 5; Length 101241;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 AATGTAACCAAAATCAACTGA 118  
|||||  
Db 39751 AATGTAACCAAAATCAACTGA 39729  
|||||

**RESULT 13** AC107028 126716 bp DNA linear PRI 29-MAR-2003  
**LOCUS** Homo sapiens 3 BAC RP11-547K2 (Rosewell Park Cancer Institute Human BAC library) complete sequence.  
**ACCESSION** AC107028  
**VERSION** AC107028.5 GI:28006935

**KEYWORDS** HTG.

**SOURCE** Homo sapiens (human)

**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**REFERENCE** 1 (bases 1 to 126716)

**AUTHORS** Alzbrooks, S. L., Amaral, H. C., Adio-Oduola, B., Ali-Osman, F. R., Allen, C., Muzny, D. M., Adams, C., Datta, A., Aye, J. R., Ayele, M., Banks, T., Barbieri, J., Benton, J., Bimberg, K., Blankenburg, K., Bonini, N. P., Bouck, J., Bowler, S., Brice, M., Brown, B., Brown, M., Bryant, N. P., Buha, C., Burck, P., Burkett, C., Butelli, K. L., Byrd, N. C., Caron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denu, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durkin, K. J., Barnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falle, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichthage, O., Lien, C., Liu, J., Liu, W., Louiege, R., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabac, K., Moore, S., Morgan, M., Morris, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, D., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scheer, S., Scott, G., Shen, H., Shooshitari, N., Sison, I., Sodergren, E., Sonalke, F., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svetek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vazquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Naylor, S. L., Weinstein, G. and Gibbs, R.

**TITLE** Direct Submission

**JOURNAL** Unpublished

**REFERENCE** 2 (bases 1 to 126716)

**AUTHORS** Worley, K. C.

**JOURNAL** Direct Submission

Submitted (14-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**REFERENCE** 3 (bases 1 to 126716)

**AUTHORS** Worley, K. C.

**JOURNAL** Direct Submission

Submitted (29-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**REFERENCE** 4 (bases 1 to 126716)

**AUTHORS** Worley, K. C.

**JOURNAL** Direct Submission

Submitted (30-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**REFERENCE** 5 (bases 1 to 126716)

**AUTHORS** Worley, K. C.

**JOURNAL** Direct Submission

Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 6 (bases 1 to 126716)  
 AUTHORS Morley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 7 (bases 1 to 126716)  
 AUTHORS Morley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JUN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 8 (bases 1 to 126716)  
 AUTHORS Morley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 9 (bases 1 to 126716)  
 AUTHORS Morley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jan 29, 2003 this sequence version replaced gi:20340431.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES  
 source Location/Qualifiers  
 1. 126716  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /clone="RP11-547K2"  
 complement(1..3011)  
 /note="overlaps bases 1..3011 of clone AC026305"  
 /function="clone overlap"  
 repeat\_region complement(28..209)

repeat\_region /rpt\_family="MER34"  
 219..973  
 /rpt\_family="LIM4"  
 repeat\_region 1409..1455  
 /rpt\_family="AluY"  
 repeat\_region 1456..1754  
 /rpt\_family="AluY"  
 repeat\_region 1762..1998  
 /rpt\_family="AluSC"  
 repeat\_region 2145..2195  
 /rpt\_family="AT-rich"  
 repeat\_region 2612..2921  
 /rpt\_family="AluSX"  
 repeat\_region 3493..3520  
 /rpt\_family="AT-rich"  
 repeat\_region 3573..3598  
 /rpt\_family="AT-rich"  
 repeat\_region 5930..5955  
 /rpt\_family="AT-rich"  
 repeat\_region complement(6168..6478)  
 /rpt\_family="AluSP"  
 repeat\_region 6945..6984  
 /rpt\_family="TTTA)n"  
 repeat\_region complement(6986..7276)  
 /rpt\_family="AluSX"  
 repeat\_region 11796..11975  
 /rpt\_family="MERSA"  
 repeat\_region complement(13903..13967)  
 /rpt\_family="L2"  
 repeat\_region 14192..14235  
 /rpt\_family="MIR"  
 repeat\_region 16052..16078  
 /rpt\_family="AT-rich"  
 repeat\_region complement(17082..17461)  
 /rpt\_family="MSTD"  
 repeat\_region complement(17583..17934)  
 /rpt\_family="MER21B"  
 repeat\_region 17944..17997  
 /rpt\_family="TTATA)n"  
 repeat\_region complement(18003..18257)  
 /rpt\_family="AluB"  
 repeat\_region complement(18261..18557)  
 /rpt\_family="MER21B"  
 repeat\_region 19148..19215  
 /rpt\_family="AT-rich"  
 repeat\_region 19234..19342  
 /rpt\_family="FLAM\_C"  
 repeat\_region 19377..19428  
 /rpt\_family="A-rich"

Query Match 0.7%; Score 23; DB 9; Length 126716;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1673 TCTTTACTCTTAAATTTTCT 1695  
 Db 47492 TCTTTACTCTTAAATTTTCT 47470

RESULT 14  
 BX005128/c 152345 bp DNA linear VRT 04-OCT-2003  
 LOCUS Zebrafish DNA sequence from clone CH211-129H14 in linkage group 17,  
 DEFINITION complete sequence.  
 ACCESSION BX005128  
 VERSION BX005128.4 GI:37518236  
 KEYWORDS HTG.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 152345)

AUTHORS  
TITLE

## COMMENT

Sehra, H.  
Direct Submission  
Submitted (03-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Oct 5, 2003 this sequence version replaced gi:30141705.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zf1sh-help@sanger.ac.uk  
-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:

Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Repeat names  
beginning 'dr' were identified by the Recon repeat discovery system  
(Zhifeng Bao and Sean Eddy, submitted), and those beginning 'ctr'  
were identified by Rick Waterman (Stephen Johnson lab, WashU). For  
further information see

[http://www.sanger.ac.uk/Projects/D\\_reiro/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_reiro/fishmask.shtml)  
CH211-129H14 is from a CHORI-211 BAC library

VECTOR: PTARBAC2.1  
Clone-derived Zebrafish pUC subclones occasionally display  
inconsistency over the length of mononucleotide A/T runs and  
conserved TA repeats. Where this is found the longest good quality  
representation will be submitted.

FEATURES  
source

1. 152345  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-129H14"  
/clone\_lib="CHORI-211"

## ORIGIN

## Query Match

Best Local Similarity 0.7%; Score 23; DB 5; Length 152345;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3141 TTAGTATTTCCTTTTCTTTTCTTGT 3163  
|||||  
2054 TTAGTATTTCCTTTTCTTTTCTTGT 2032

RESULT 15  
AC048381

LOCUS AC048381 180925 bp DNA linear HTG 24-AUG-2002  
DEFINITION Homo sapiens chromosome 3 clone RP11-547K2 map 3, WORKING DRAFT  
SEQUENCE, 37 unordered pieces.  
AC048381

ACCESSION AC048381.3 GI:8084282  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

## REFERENCE

1 (bases 1 to 180925)

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Unpublished  
2 (bases 1 to 180925)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Baettien, V., Beda, F.,  
Bogunlavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Hayes, B., Heatford, A., Horton, L.,  
Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L.,  
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lebecky, J.,  
Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McEwan, K., McPherson, R.,  
Meldrum, J., Menus, L., Miho, T., Miranda, C., Mieng, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,  
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 180925)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Baettien, V., Beda, F.,  
Bogunlavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Hayes, B., Heatford, A., Horton, L.,  
Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L.,  
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lebecky, J.,  
Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McEwan, K., McPherson, R.,  
Meldrum, J., Menus, L., Miho, T., Miranda, C., Mieng, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,  
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

## COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7770638.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@genome.wi.mit.edu

Center Project Information  
Center project name: L9685  
Center clone name: 547\_K\_2

Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 158029 bases at least Q40  
Consensus quality: 168695 bases at least Q30  
Consensus quality: 173285 bases at least Q20  
Insert size: 183000; agarose-fp

Insert size: 177325; sum-of-ctnigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality coverage: 3.6 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently consists of 37 ctnigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the ctnigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1010: ctnig of 1010 bp in length
1011 1110: gap of 100 bp
1111 2487: ctnig of 1377 bp in length
2488 2587: gap of 100 bp
2588 3720: ctnig of 1133 bp in length
3721 3820: gap of 100 bp
3821 4892: ctnig of 1072 bp in length
4893 4992: gap of 100 bp
4993 5489: ctnig of 497 bp in length
5490 5589: gap of 100 bp
5590 7151: ctnig of 1562 bp in length
7152 7251: gap of 100 bp
7252 8655: ctnig of 1404 bp in length
8656 8755: gap of 100 bp
8756 10134: ctnig of 1379 bp in length
10135 10235: gap of 100 bp
10236 13107: ctnig of 2873 bp in length
13108 13207: gap of 100 bp
13208 15265: ctnig of 2058 bp in length
15266 15365: gap of 100 bp
15366 17768: ctnig of 2403 bp in length
17769 17868: gap of 100 bp
17869 20233: ctnig of 2365 bp in length
20234 20333: gap of 100 bp
20334 22968: ctnig of 2635 bp in length
22969 23068: gap of 100 bp
23069 25333: ctnig of 2265 bp in length
25334 25433: gap of 100 bp
25434 28978: ctnig of 3545 bp in length
28979 29078: gap of 100 bp
29079 31604: ctnig of 2526 bp in length
31605 31704: gap of 100 bp
31705 34400: ctnig of 2696 bp in length
34401 34501: gap of 100 bp
34502 37255: ctnig of 2755 bp in length
37256 37355: gap of 100 bp
37356 40057: ctnig of 2702 bp in length
40058 40157: gap of 100 bp
40158 44190: ctnig of 4033 bp in length
44191 44290: gap of 100 bp
44291 46785: ctnig of 2495 bp in length
46786 46885: gap of 100 bp
46886 52597: ctnig of 5712 bp in length
52598 52697: gap of 100 bp
52698 59026: ctnig of 6329 bp in length
59027 59126: gap of 100 bp
59127 64431: ctnig of 5305 bp in length
64432 64531: gap of 100 bp
64532 69115: ctnig of 4584 bp in length
69116 69215: gap of 100 bp
69216 73866: ctnig of 4651 bp in length
73867 73966: gap of 100 bp
73967 79516: ctnig of 5550 bp in length
79517 79616: gap of 100 bp
79617 85688: ctnig of 6072 bp in length
85689 92058: ctnig of 6270 bp in length
92059 92158: gap of 100 bp
92159 99039: ctnig of 6881 bp in length
99040 99139: gap of 100 bp
108885: ctnig of 9746 bp in length

```

```

* 108886 108985: gap of 100 bp
* 108986 116971: ctnig of 7986 bp in length
* 116972 117071: gap of 100 bp
* 117072 126718: ctnig of 9647 bp in length
* 126719 126818: gap of 100 bp
* 126819 137354: ctnig of 10536 bp in length
* 137355 137455: gap of 100 bp
* 137455 146943: ctnig of 9489 bp in length
* 146944 147043: gap of 100 bp
* 147044 157785: ctnig of 10742 bp in length
* 157786 157885: gap of 100 bp
* 157886 180925: ctnig of 23040 bp in length.
FEATURES
source
1. 180925
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-547K2"
/clone_11b="RPCT-11 Human Male BAC"
1. 1010
/note="assembly_fragment"
1111. 2487
/note="assembly_fragment"
2588. 3720
/note="assembly_fragment"
3821. 4892
/note="assembly_fragment"
4993. 5489
/note="assembly_fragment"
clone_end:T7
vector_side:right"
5590. 7151
/note="assembly_fragment"
7252. 8655
/note="assembly_fragment"
8756. 10134
/note="assembly_fragment"
10235. 13107
/note="assembly_fragment"
13208. 15265
/note="assembly_fragment"

```

```

Query Match 0.7%; Score 23; DB 2; Length 180925;
Best Local Similarity 100.0%; Pred.No.1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1673 TCTTTACTCTTAATTTTCT 1695
|||||
DB 16692 TCTTTACTCTTAATTTTCT 16714
|||||

```

Search completed: March 30, 2004, 14:48:22  
Job time : 12264.4 secs

**This Page Blank (uspto)**